

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 14:53:14 ; Search time 5104.38 Seconds
(without alignments)
9184.983 Million cell updates/sec

Title: US-09-989-981A-9_COPY_3436_5005
Perfect score: 1570
Sequence: 1 cgaagcatcctgaagtacag.....ctagagagcaaaccagagc 1570

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query						Description
	No.	Score	Match	Length	DB	ID	
c	1	206.4	13.1	306	10	BB605863	BB605863 BB605863
	2	166.4	10.6	339	10	BB869579	BB869579 BB869579
c	3	164.4	10.5	303	10	BB870338	BB870338 BB870338
	4	164.4	10.5	658	13	BY742680	BY742680 BY742680
	5	164.4	10.5	713	10	BB598373	BB598373 BB598373
c	6	143.4	9.1	663	28	AZ051299	AZ051299 sito0006
	7	137.2	8.7	460	9	AA239884	AA239884 mx81d01.r
	8	133.4	8.5	502	9	AA237916	AA237916 mx14e08.r
c	9	130	8.3	583	13	BY705076	BY705076 BY705076
c	10	130	8.3	3623	11	AK004871	AK004871 Mus muscu
c	11	123	7.8	500	9	AI151811	AI151811 ui46c10.y
c	12	123	7.8	510	10	BB610072	BB610072 BB610072
c	13	115	7.3	511	9	AI157365	AI157365 ui45h01.y
c	14	112.8	7.2	599	28	BH321870	BH321870 CH230-7C1
	15	103.4	6.6	764	12	BI246567	BI246567 602958477
c	16	64	4.1	435	9	AI574075	AI574075 uj67h11.y
	17	59.4	3.8	606	14	CD502116	CD502116 CDA54-H04
c	18	54	3.4	1201	13	BX381961	BX381961 BX381961
	19	49.2	3.1	1201	13	BX381961	BX381961 BX381961
	20	48.2	3.1	925	29	CNS03RDA	AL257095 Tetraodon
c	21	46	2.9	1201	13	BX376097	BX376097 BX376097
	22	43.8	2.8	432	13	BY252099	BY252099 BY252099
c	23	43	2.7	294	14	CB424734	CB424734 599034 MA
	24	42.4	2.7	997	29	CNS005TE	AL060767 Drosophil
c	25	40.6	2.6	629	9	AA543856	AA543856 vk34a07.r
	26	39.8	2.5	294	14	CB417759	CB417759 590490 MA
c	27	39.8	2.5	455	28	AQ977239	AQ977239 RPCI-23-3
c	28	39.6	2.5	1201	13	BX399635	BX399635 BX399635
c	29	39.4	2.5	662	10	BB612448	BB612448 BB612448
	30	39.4	2.5	1141	13	BX414498	BX414498 BX414498
	31	39.2	2.5	361	12	BI028780	BI028780 CM0-MT018
c	32	39.2	2.5	837	14	CK203027	CK203027 FGAS01155
c	33	39	2.5	458	9	AA914287	AA914287 vy99b08.r
c	34	39	2.5	493	9	AA060852	AA060852 mj86d02.r
c	35	39	2.5	525	9	AA882149	AA882149 vx38e02.r
c	36	39	2.5	856	12	BI416074	BI416074 602987346
c	37	39	2.5	929	14	CB590318	CB590318 AGENCOURT
c	38	39	2.5	933	11	AK010167	AK010167 Mus muscu
c	39	39	2.5	946	11	BC032992	BC032992 Mus muscu
c	40	39	2.5	1101	29	CNS0037Q	AL064465 Drosophil
	41	38.8	2.5	834	13	BQ752298	BQ752298 EST632861
c	42	38.6	2.5	551	9	AA390068	AA390068 mv35b05.r
	43	38.6	2.5	987	29	CNS00418	AL066537 Drosophil
c	44	38.6	2.5	1201	13	BX335650	BX335650 BX335650
	45	38.4	2.4	407	9	AA525033	AA525033 nh36c06.s

46	38.4	2.4	412	9	AA524916	AA524916 nh31a09.s
47	38.4	2.4	472	14	CF486702	CF486702 POL1_39_F
48	38.4	2.4	910	29	CNS0060N	AL065629 Drosophil
c 49	38.2	2.4	472	28	BZ850575	BZ850575 CH240_280
c 50	38.2	2.4	560	28	BZ849786	BZ849786 CH240_280

ALIGNMENTS

RESULT 1
BB605863/c

LOCUS BB605863 306 bp mRNA linear EST 05-DEC-2000

DEFINITION BB605863 RIKEN full-length enriched, 0 day neonate lung Mus musculus cDNA clone E030013I04 5', mRNA sequence.

ACCESSION BB605863

VERSION BB605863.1 GI:11557265

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 306)

AUTHORS Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Aizawa,K. et al. 2000)

JOURNAL Unpublished (2000)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

VERSION BB869579.1 GI:17115789
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 339)
 AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
 Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
 Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
 Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.
 TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. . 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
 Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
 and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
 Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.
 FEATURES Location/Qualifiers
 source 1. .339
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G630014E22"
 /sex="male"
 /tissue_type="intestinal mucosa"

/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male
intestinal mucosa"

ORIGIN

Query Match 10.6%; Score 166.4; DB 10; Length 339;
Best Local Similarity 96.6%; Pred. No. 3.7e-38;
Matches 170; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 403 GCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAG 462
|||||
Db 2 GCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAG 61

Qy 463 GGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAG 522
|||||
Db 62 GGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAG 121

Qy 523 AGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGAC 578
||||| | | | |
Db 122 AGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTC 177

RESULT 3

BB870338/c

LOCUS BB870338 303 bp mRNA linear EST 27-NOV-2001

DEFINITION BB870338 RIKEN full-length enriched, adult male intestinal mucosa
Mus musculus cDNA clone G630020H06 5', mRNA sequence.

ACCESSION BB870338

VERSION BB870338.1 GI:17116548

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 303)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 658)
 AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
 Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
 Rogers,J., Birney,E. and Hayashizaki,Y.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 22354683
 PUBMED 12466851
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
 Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
 Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
 Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
 Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
 Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
 Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1. .658

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="C730040P06"

/sex="male"

/tissue_type="liver tumor"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, adult male liver tumor"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCCC 3']. cDNA

was cleaved with BamHI and XhoI. Vector: a modified

pBluescript KS(+) after bulk excision from Lambda FLC I.

Tissue was provided by William A. Held, Roswell Park

Cancer Institute, Department of Molecular and Cellular

Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose

assistance we gratefully acknowledge."

ORIGIN

Query Match 10.5%; Score 164.4; DB 13; Length 658;

Best Local Similarity 96.6%; Pred. No. 2.2e-37;

Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 405 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 464
 |||
 Db 1 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 60

Qy 465 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 524
 |||
 Db 61 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 120

Qy 525 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGAC 578
 |||
 Db 121 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTC 174

RESULT 5

BB598373

LOCUS BB598373 713 bp mRNA linear EST 26-OCT-2001

DEFINITION BB598373 RIKEN full-length enriched, adult male liver tumor Mus musculus cDNA clone C730003G04 5', mRNA sequence.

ACCESSION BB598373

VERSION BB598373.2 GI:16450340

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 713)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT On Dec 1, 2000 this sequence version replaced gi:11506974.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)

cdna library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

1. .713

```
/organism="Mus musculus"
```

```
/mol type="mRNA"
```

```
/db xref="taxon:10090"
```

```
/clone="C730003G04"
```

```
/sex="male"
```

```
/tissue type="liver tumor"
```

```
/dev stage="adult"
```

```
/lab host="DH10B"
```

```
/clone_lib="RIKEN full-length enriched, adult male liver
tumor"
```

/note="Site 1: SalI; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTTVN 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCCC 3'] . cDNA

was cleaved with BamHI and XhoI. Vector: a modified

pBluescript KS(+) after bulk excision from Lambda FLC I.

Tissue was provided by William A. Held, Roswell Park

Cancer Institute, Department of Molecular and Cellular

Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose

assistance we gratefully acknowledge."

Query Match 10.5%; Score 164.4; DB 10; Length 713;

Best Local Similarity 96.6%; Pred. No. 2.4e-37;

Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 405 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 464

|||||

Db 1 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 60

Qy 465 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 524
 |||
 Db 61 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 120

Qy 525 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTTACAGCGTCAGGTAAGGGGAC 578
 |||
 Db 121 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTTACAGCGTCAGCAACCGTGTC 174

RESULT 6

AZ051299/c

LOCUS AZ051299 663 bp DNA linear GSS 28-MAR-2001

DEFINITION sito0006 Human Homo sapiens genomic clone CITB-978SK-B 569J16 T7
 end, genomic survey sequence.

ACCESSION AZ051299

VERSION AZ051299.1 GI:13470256

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 663)

AUTHORS Lu, K., Lee, M. and Patel, S.B.

TITLE High-Resolution Physical and Transcript Map of Human Chromosome
 2p21 Containing the Sitosterolemia Locus

JOURNAL Unpublished (2000)

COMMENT Contact: Patel SB

Division of Endocrinology, Diabetes and Medical Genetics

Medical University of South Carolina

Strom Thurmond Bldg., Room 541, 114 Doughty Street, Charleston, SC
 29403, USA

Tel: 843 876 5227

Fax: 843 876 5133

Email: patelsb@musc.edu

Seq primer: T7

Class: BAC ends.

FEATURES Location/Qualifiers

source

1. .663

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="CITB-978SK-B 569J16"

/clone_lib="Human"

ORIGIN

Query Match 9.1%; Score 143.4; DB 28; Length 663;

Best Local Similarity 80.4%; Pred. No. 4.5e-31;

Matches 168; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1082 GGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGTC 1141
 || || | |||| | ||| ||| ||||| ||| | |||||
 Db 211 GGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCTTCCAGGGCCCCAAGCTGCCTTTGTC 152

Qy 1142 CCTTGTGGGATTTCTTTAAAGCAACCGTGTCGGGCCTTGGTGGAACATCAAATCATGCC 1201
 | | ||| ||||| |||| | ||| ||||| ||||| ||| |||
 Db 151 CTCCTGGGGTTTCTTTAAAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCC 92


```

Qy      1202 AGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGA 1261
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      91   GGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGA 32

Qy      1262 TTATGTGCATCTTAGGCAGCTCAGGTAAG 1290
          | ||||| ||||| ||||| ||||| |||||
Db      31   TCATGTGCATCCTAGGAAGCTCAGGTAAG 3

```

AA239884

LOCUS	AA239884	460 bp	mRNA	linear	EST 03-MAR-1997
-------	----------	--------	------	--------	-----------------

DEFINITION mx8ld01.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:692737 5' similar to WP:F19B6.4 CE05669 WHITE PROTEIN LIKE ;, mRNA sequence.

ACCESSION AA239884

VERSION AA239884.1 GI:1863923

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM	Mus musculus
----------	--------------

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 460)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:426297

Possible reversed clone: similarity on wrong strand

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 413.

FEATURES

source

Location/Qualifiers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

1. .460

```
/organism="Mus musculus"
```

```
/mol type="mRNA"
```

```
/db xref="taxon:10090"
```

```
/clone="IMAGE:692737"
```

```
/tissue type="Liver"
```

```
/lab host="DH10B"
```

```
/clone lib="Soares mouse NML"
```

```
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand
```

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 8.7%; Score 137.2; DB 9; Length 460;
Best Local Similarity 97.9%; Pred. No. 2.6e-29;
Matches 139; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 427 GTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTC 486
|||||
Db 1 GTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTC 60

Qy 487 TGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCC 546
|||||
Db 61 TGAGCTCCCTGGACATAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCC 120

Qy 547 TGCATGTGTCCTACAGCGTCAG 568
|||||
Db 121 TGCATGTGTCCTACAGCGTCAG 142

RESULT 8

AA237916

LOCUS AA237916 502 bp mRNA linear EST 03-MAR-1997

DEFINITION mx14e08.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:680198 5' similar to SW:BROW_DROME P12428 BROWN PROTEIN. ;, mRNA sequence.

ACCESSION AA237916

VERSION AA237916.1 GI:1861938

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 502)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:419902

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 459.

FEATURES Location/Qualifiers

source 1. .502

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:680198"
/tissue_type="Liver"
/lab_host="DH10B"
/clone_lib="Soares mouse NML"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

ORIGIN

```

Query Match          8.5%;  Score 133.4;  DB 9;  Length 502;
Best Local Similarity 95.8%;  Pred. No. 3.8e-28;
Matches 137;  Conservative 0;  Mismatches 6;  Indels 0;  Gaps 0;

Qy      436 CCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCC 495
          |||
Db      2   CCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCC 61

Qy      496 TGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGT 555
          |||
Db      62  TGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGT 121

Qy      556 CCTACAGCGTCAGGTAAGGGGAC 578
          |||
Db      122 CCTACAGCGTCACGAACCGTGTC 144

```

RESULT 9

BY705076/c

LOCUS BY705076 583 bp mRNA linear EST 16-DEC-2002

DEFINITION BY705076 RIKEN full-length enriched, adult male liver Mus musculus
cDNA clone 1300003C16 5', mRNA sequence.

ACCESSION BY705076

VERSION BY705076.1 GI:27116215

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 583)

AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,

Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:<http://genome.gsc.riken.go.jp/>
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

FEATURES Location/Qualifiers
 source 1. .583
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1300003C16"
 /sex="male"
 /tissue_type="liver"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult male liver"

ORIGIN

Query Match 8.3%; Score 130; DB 13; Length 583;
Best Local Similarity 100.0%; Pred. No. 4.3e-27;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
 |||
Db 131 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 72

Qy 61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
 |||
Db 71 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 12

Qy 121 TTTGCTCCTT 130
 |||
Db 11 TTTGCTCCTT 2

RESULT 10
AK004871/c

LOCUS AK004871 3623 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
 library, clone:1300003C16 product:ATP-BINDING CASSETTE, SUB-FAMILY
 G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full insert
 sequence.

ACCESSION AK004871

VERSION AK004871.1 GI:12836380

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

 ORGANISM Mus musculus

 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

 AUTHORS Carninci,P. and Hayashizaki,Y.

 TITLE High-efficiency full-length cDNA cloning

 JOURNAL Meth. Enzymol. 303, 19-44 (1999)

 MEDLINE 99279253

 PUBMED 10349636

REFERENCE 2

 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes

 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 3623)
 AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
 Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
 Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
 Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
 Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
 Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
 Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
 Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
 Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
 Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)
 COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGAGAGAGAGCGCCGCAACTCGAGTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second

strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

```

FEATURES             Location/Qualifiers
    source             1..3623
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="FANTOM DB:1300003C16"
                        /db_xref="MGI:1896857"
                        /db_xref="taxon:10090"
                        /clone="1300003C16"
                        /sex="male"
                        /tissue_type="liver"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="adult"
    CDS                69..2090
                        /note="unnamed protein product; ATP-BINDING CASSETTE,
                        SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus]
                        (SWISSPROT|Q9DBM0, evidence: FASTY, 92%ID, 96.7%length,
                        match=1796)
                        putative"
                        /codon_start=1
                        /protein_id="BAB23630.1"
                        /db_xref="GI:12836381"
                        /translation="MAEKTKEETQLWNGTVLQDASQGLQDSLFSSES DNSLYFTYSQQ
                        SNTLEVRDLTYQVDIASQVPWF EQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQOML
                        AIIGSSGCGRASLLDVITGRGHGGMKSGQI WINGQFPSTPQLVRKCVAHVRQHDQLLP
                        NLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGE
                        RRRVSIQVQLLWNPGILILDEPTSGLDSFTAHNLVTTLSRLAKGNRLVLISLHQPRSD
                        IFRLEDLVLLMTSGTPIYLGAAQQM VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKE
                        REVATVEKAQSLAALFLEKVGQFDDFLWKAEAKEINTSTHTVSLTLTQDTCGTAVEL
                        PGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACIMSLIIGFLYYGHGAKQLSFMDTAA
                        LLEMIGALIPFNVILDVVS KCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYV
                        ILYAMPIYWLTNLRPVPELFL LHFLLVVLVVFCCRTMALAASAMLPTFHMSSFFCNAL
                        YNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSI
                        LGDTMISAMDLSNHPYLA IYLVIGISYGFLFLYYLSLKLKIKQKSIQDW"
    polyA_signal       3605..3610
                        /note="putative"
    polyA_site         3623
                        /note="putative"

```

ORIGIN

Query Match 8.3%; Score 130; DB 11; Length 3623;
 Best Local Similarity 100.0%; Pred. No. 1.3e-26;
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
        |||
Db      131 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 72

Qy      61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
        |||
Db      71 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 12

Qy      121 TTTGCTCCTT 130

```

Db |||||||
11 TTTGCTCCTT 2

RESULT 11
AI151811/c
LOCUS AI151811 500 bp mRNA linear EST 30-SEP-1998
DEFINITION ui46c10.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1885458 5', mRNA sequence.
ACCESSION AI151811
VERSION AI151811.1 GI:3680280
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 500)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:969782
Seq primer: custom primer used
High quality sequence stop: 499.
FEATURES Location/Qualifiers
source 1. .500
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1885458"
 /dev_stage="embryo, 14 dpc"
 /lab_host="DH10B"
 /clone_lib="Sugano mouse embryo mewa"
 /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TGTTGGCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer

ORIGIN

```

Qy          1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
            |||
Db          123 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 64

Qy          61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
            |||
Db          63 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 4

Qy          121 TTT 123
            |||
Db          3 TTT 1

```

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:<http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* . 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES
source Location/Qualifiers
1. .510
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1300007N20"
/sex="male"
/tissue_type="liver"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male liver"

ORIGIN

Query Match 7.8%; Score 123; DB 10; Length 510;
Best Local Similarity 100.0%; Pred. No. 5e-25;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
|||||
Db 126 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 67
Qy 61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
|||||
Db 66 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 7
Qy 121 TTT 123
|||
Db 6 TTT 4

RESULT 13
AI157365/c

LOCUS AI157365 511 bp mRNA linear EST 30-SEP-1998
DEFINITION ui45h01.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1885393 5', mRNA sequence.
ACCESSION AI157365
VERSION AI157365.1 GI:3685834

KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 511)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:969717
 Seq primer: custom primer used
 High quality sequence stop: 480.

FEATURES Location/Qualifiers
 source 1. .511
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1885393"
 /dev_stage="embryo, 14 dpc"
 /lab_host="DH10B"
 /clone_lib="Sugano mouse embryo mewa"
 /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
 Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
 with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTTT];
 double-stranded cDNA was ligated to a DraIII adaptor
 [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
 sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
 CACCATGTG). XhoI should be used to isolate the cDNA
 insert. Size selection was performed to exclude fragments
 <1.5kb. Library constructed by Dr. Sumio Sugano
 (University of Tokyo Institute of Medical Science).
 Custom primers for sequencing: 5' end primer
 CTTCTGCTCTAAAAGCTGCG and 3' end primer
 CGACCTGCAGCTCGAGCACA."

ORIGIN

Query Match 7.3%; Score 115; DB 9; Length 511;
 Best Local Similarity 100.0%; Pred. No. 1.3e-22;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 116 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 57

Qy 61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCT 115
 |||
 Db 56 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCT 2

RESULT 14
 BH321870/c

LOCUS BH321870 599 bp DNA linear GSS 03-DEC-2001
 DEFINITION CH230-7C13.TVB CHORI-230 Segment 1 Rattus norvegicus genomic clone
 CH230-7C13, genomic survey sequence.

ACCESSION BH321870
 VERSION BH321870.1 GI:17252584
 KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 599)

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
 Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D.,
 Riggs, F., de Jong, P. and Fraser, C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: CH230-7C13.TJB
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/orering_information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 7 row: C column: 13
 Seq primer: T7
 Class: BAC ends.

FEATURES Location/Qualifiers

source 1. .599
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SsNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-7C13"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 1"
 /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 7.2%; Score 112.8; DB 28; Length 599;
 Best Local Similarity 73.4%; Pred. No. 6.3e-22;
 Matches 160; Conservative 1; Mismatches 48; Indels 9; Gaps 1;

```

Qy      1352 CTAAGCACAATGTTTAAGAAGTRAGTTTAAGTTGTAGAGAGGCAGCCATGCATTTGGCAT 1411
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      599 CCAAAAAAAAAAAAAAAAAAGAAATGAGTTTAAGTTGGAGAGAAAAGGCTATGCATTTAGCAT 540

Qy      1412 TTGAATACAATCTGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCT 1471
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      539 TTGAACAAAATCTAGTGA-----CTGTGAATAGAACCTGGTATCAAAGTGAAACCT 489

Qy      1472 TGAGGAAAATCCCTGGAAAGAGTGGAAAGTCCTGCCTAACACGTAAGTGCCTTCTTTGCT 1531
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      488 TAGGAAAAGTATCTGGAGAAAGTGGGAAGTCCTGCCTGACGTGTAAGGACTTTCTGTGCT 429

Qy      1532 TGTTTGATTGACTGTGATGCTAGAGAGCAAACCCAGAG 1569
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      428 TGTTTGATTGACTGTGGTGCTGGAGATCAGAGCCTCAG 391
  
```

RESULT 15

BI246567

LOCUS BI246567 764 bp mRNA linear EST 17-JUL-2001

DEFINITION 602958477F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5124187 5', mRNA sequence.

ACCESSION BI246567

VERSION BI246567.1 GI:14790652

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 764)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11303 row: i column: 20

High quality sequence start: 2

High quality sequence stop: 666.

FEATURES

source

Location/Qualifiers

1. .764

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5124187"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NCI_CGAP_Li9"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 6.6%; Score 103.4; DB 12; Length 764;
Best Local Similarity 98.3%; Pred. No. 4.9e-19;
Matches 115; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1162 AGCAACCGTGTCTGGGCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAA 1221
|||||
Db 141 AGCAACCGTGTCTGGGCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAA 200

Qy 1222 ATCCTCAAAGATGTCTCCTTG-TACATCGAGAGTGGCCAGATTATGTGCATCTTAGG 1277
|||||
Db 201 ATCCTCAAAGATGTCTCCTTGATACATCGAGAGTGGCCAGATTATGTGCATCTTACG 257

RESULT 16

AI574075/c

LOCUS AI574075 435 bp mRNA linear EST 29-MAR-1999

DEFINITION uj67h11.y1 Sugano mouse liver mlia Mus musculus cDNA clone
IMAGE:1925061 5', mRNA sequence.

ACCESSION AI574075

VERSION AI574075.1 GI:4537449

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 435)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:981353

Seq primer: custom primer used

High quality sequence stop: 432.

FEATURES

source

Location/Qualifiers

1. .435

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL"

/db_xref="taxon:10090"

```

/clone="IMAGE:1925061"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse liver mlia"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGAGCACA."

```

ORIGIN

```

Query Match          4.1%; Score 64; DB 9; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
        |||
Db      64 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 5

Qy      61 CATG 64
        |||
Db      4 CATG 1

```

RESULT 17

CD502116

LOCUS CD502116 606 bp mRNA linear EST 12-JUN-2003

DEFINITION CDA54-H04.xld-t SHGC-CDA Gasterosteus aculeatus cDNA clone
CDA54-H04 5', mRNA sequence.

ACCESSION CD502116

VERSION CD502116.1 GI:31429142

KEYWORDS EST.

SOURCE Gasterosteus aculeatus (three spined stickleback)

ORGANISM Gasterosteus aculeatus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.

REFERENCE 1 (bases 1 to 606)

AUTHORS Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.

TITLE Expressed sequence tags from Gasterosteus aculeatus

JOURNAL Unpublished (2003)

COMMENT Contact: Kingsley, DM

HHMI and Department of Developmental Biology

Stanford University School of Medicine

Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA

Tel: 650 725 5954

Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 54
High quality sequence stop: 606.

FEATURES Location/Qualifiers
 source 1. .606
 /organism="Gasterosteus aculeatus"
 /mol_type="mRNA"
 /strain="Salinas river, CA"
 /db_xref="taxon:69293"
 /clone="CDA54-H04"
 /sex="mixed male and female"
 /tissue_type="heads and internal organs combined"
 /dev_stage="adult"
 /clone_lib="SHGC-CDA"
 /note="Vector: lambda ZAP Express/pBK-CMV; Site_1: EcoR1
 (5' adaptor); Site_2: XhoI (3' linker primer); The mixed
 organ cDNA library was generated using the ZAP-cDNA method
 by Stratagene. First strand cDNA synthesis was primed with
 a a 50 bp linker primer containing an oligo dT sequence
 preceeded by a synthetic XhoI site. 5 prime adaptors were
 used containing an EcoR1 cohesive end. The finished cDNAs
 were inserted in to the ZAP express vector
 unidirectionally in the sense orientation with respect to
 the lacZ promoter of pBK-CMV. An amplified library was
 prepared from approximately 3 million primary clones in
 the lambda ZAP Express vector. In vivo excision was then
 used to generate individual pBK-CMV phagemid clones for
 EST sequencing."

ORIGIN

Query Match 3.8%; Score 59.4; DB 14; Length 606;
Best Local Similarity 67.2%; Pred. No. 6.6e-06;
Matches 84; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy      1162 AGCAACCGTGTCTGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAA 1221
          || | ||||| || || ||||| || | || | | || | || | ||
Db      277 AGTGAGCGTGTGGGTCCGTGGTGGGACTTACCCTCCTTCAGGAAGCGATGGACTCGTCAG 336

Qy      1222 ATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGC 1281
          ||||| ||||| ||||| || | || | || | ||||| || |||| | ||||
Db      337 ATCCTCAATGATGTCTCCTTCCACGTGGACAGCGGGCAGATCATGGGCATACTGGGCAAT 396

Qy      1282 TCAGG 1286
          |||||
Db      397 TCAGG 401
```

RESULT 18

BX381961/c

LOCUS BX381961 1201 bp mRNA linear EST 08-MAY-2003

DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 clone CS0DI072YF05 3-PRIME, mRNA sequence.

ACCESSION BX381961

VERSION BX381961.1 GI:30453007

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1201)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> InVitroGen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DI072CC03NP1.

FEATURES Location/Qualifiers

source 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DI072YF05"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 3.4%; Score 54; DB 13; Length 1201;
 Best Local Similarity 6.8%; Pred. No. 0.00042;
 Matches 45; Conservative 242; Mismatches 375; Indels 0; Gaps 0;

Qy 450 GAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCG 509
 | |:| | | : | || |:: ::::|: |: :: |: :: :|:: : :
 Db 1192 GSGKGGGGCCSCGCCCCCMMMVMGMMGMKNGGMMGMMMMMMMMGMMMMMBGBM 1133

Qy 510 GTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTACAGCGTCAGG 569
 :: : | :: |::| : || : |::| : : |::|:::
 Db 1132 MVGKGVGKGGKGGKGMVMMGGCMSCVKKGGSNKGCBGCMGKGCCKGCKGCMGKGVVKM 1073

Qy 570 TAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGGGT 629
 : :: : ::::| :::: : :|: : : : :::: | : : : : :|::
 Db 1072 GBMMNKKKMMSKVMGMMGKMMMKBKMMGKKKKKNNKKKBBKTVKKKKTVKKKBKMTGKK 1013

Qy 630 GGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAA 689
 :: : |:: : :|:: :::: : : ::::|:: : :
 Db 1012 KKKKGKMMGGBMMKVMMGKGMVHGKGBMMBGKGMMMMTGKMVKKBMMGGKMKMKMK 953

Qy 690 CAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAACAT 749
 : : : : : :::: : : : : : : : : : : : : :| : : : : :
 Db 952 BGMVNKKTMMMBKMKNNKMMGKKVMMGKKKMMKKNMMDMMMKKGGKMMKKGNMMVK 893

Qy 750 GCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACTCG 809
 : : : : | : : : : :|: : : : : : : : : : : : : : :
 Db 892 KKNMMKKKGKKTGKKKKKKNHMMKKTMMNNKKKKMKMKKKKKKNMMVKGMKGMGKKKV 833

Qy	810	CCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAA	869
		: :: :: : :: :: : : :: : :: : :	
Db	832	MKGKGMMNMKKKMGKGBKMCVMKMMMMGKGGKMGGKMMMGMGSKMKMMNNNNNNNNNNNM	773
Qy	870	AATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAG	929
		: : : : : : : : : : : : : :	
Db	772	NNNNNMNNNNNNKNNKNAMNNMMNGNGNNKNNNNNNNMNNKMMNNANNNMMTNMMMMN	713
Qy	930	GTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCT	989
		: : : : : : : : : : : : :	
Db	712	AMNNMMNNNNKMKNNNNNNNNNNNNNNMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNKNNNMNNNNNNKM	653
Qy	990	TGGGTGGTGGCCAACCTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAA	1049
		: : : : : : : : : : : : :	
Db	652	NKNNNNMMNNMMNNNNNNNNKMNKKKKMKMMMMMNKGNMNNNNNNNNNNMMNTNNNNKNN	593
Qy	1050	AACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTC	1109
		: : : : : : : : : : : : :	
Db	592	NMNNNNMMNNNNNNNNNNNNMMNNNNNNNNNNNNNNAMGNKKNNMNKKKMNKMMNNKNTNKKM	533
Qy	1110	TG 1111	
Db	532	NG 531	

```
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

Query Match 3.1%; Score 49.2; DB 13; Length 1201;
Best Local Similarity 4.4%; Pred. No. 0.012;
Matches 32; Conservative 265; Mismatches 436; Indels 2; Gaps 1;

Qy 1298 GGGGSCSGGGGCTCC 1312
::|:| : |:|
Db 1089 CMCKGCVGCMNSCCM 1103

RESULT 20

CNS03RDA

LOCUS CNS03RDA 925 bp DNA linear GSS 01-SEP-2000

DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 049I12 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL257095

VERSION AL257095.1 GI:7978107

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.

REFERENCE 1

AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Nat. Genet. 25 (2), 235-238 (2000)

MEDLINE 20296633

PUBMED 10835645

REFERENCE 2

AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Genome Res. 10 (7), 939-949 (2000)

MEDLINE 20359837

PUBMED 10899143

REFERENCE 3 (bases 1 to 925)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

Location/Qualifiers

1. .925

/organism="Tetraodon nigroviridis"

/mol_type="genomic DNA"

/db_xref="taxon:99883"

/clone="049I12"

/clone_lib="G"

/note="Genoscope sequence ID : C0BG049BE06SP1~end :

ORIGIN

Qy 1210 TGGGACAGGCCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGCCAGATTATGTGC 1269
||| | || || || || || || || || || || || || || || || || ||
Db 7 TGGACTCGTCAGATTCTCAACGACGTCTCCTTCCACGTGGAGAGCGGCCAGATCATGGGC 66

Qy 1270 ATCTTAGGCAGCTCAGGTAAGTGCC 1294
||| | |||| |||||| | ||
Db 67 ATCCTGGGCAACTCAGGTCTGCCCC 91

BX376097/c

DEFINITION BX376097 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC022YM12 5-PRIME, mRNA sequence.

VERSION BX376097.1 GI:30434756

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1201)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2866.f

Contact : Feng Liang Email : fliang@lifetech.com URL :

<http://fulllength.invitrogen.com/> InVitroGen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DC022BG06QP1.

FEATURES	Location/Qualifiers
----------	---------------------

source

1. .1201

```
/organism="Homo sapiens"
```

```
/mol type="mRNA"
```

```
/db xref="taxon:9606"
```

```
/clone="CS0DC022YM12"
```

```
/tissue type="NEUROBLASTOMA COT 25-NORMALIZED"
```

```
/clone lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
```

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 2.9%; Score 46; DB 13; Length 1201;

Best Local Similarity 11.5%; Pred. No. 0.11;

Matches 79; Conservative 262; Mismatches 345; Indels 3; Gaps 2;

RESULT 22

LOCUS BY252099 432 bp mRNA linear EST 10-DEC-2002

ACCESSION BY252099

VERSION BY252099.1 GI:26433611
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 432)
 AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustinich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasaki,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
 Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
 Rogers,J., Birney,E. and Hayashizaki,Y.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 22354683
 PUBMED 12466851
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
 Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
 Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
 Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
 Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
 Hayashizaki,Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
source Location/Qualifiers
1. .432
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K230342H21"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match 2.8%; Score 43.8; DB 13; Length 432;
Best Local Similarity 48.4%; Pred. No. 0.26;
Matches 120; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy      544 TCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTC 603
          ||||| |  || |  |||  |  ||  |  ||  ||| ||| ||
Db      20 TCCTGCAGCTTCTCTTCCAGTTCAGCGGGTCAGGGCTGGTCGCGGAATCGCTCGGCTGTC 79

Qy      604 TGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGA 663
          | |||  ||  |||| |  |  || | |  |  ||||| | | ||
Db      80 ACCTCGCCAGCTCGTTGCCGGTGCGCTCGGGNGGGCCGTCCGTCCGTCTGTACGGAGGA 139

Qy      664 TCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGC 723
          | ||| || |  ||  || | ||| | |  |||||  ||
Db      140 TGAGGCTGGAGTGGGCGGAGGCGTGAGAACCAGTTACTTTCCTCCCAGGTGGAGCCGG 199

Qy      724 AGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGA 783
          |||  || | |  |||| |  |||| |  || ||| || | |
Db      200 AGGATTGAACTTCACCCTGAAACCACCCTCCGCCGGTCCCACCTGGCCGCCTTTACGTAA 259

Qy      784 TTTCTGCT 791
          ||| ||
Db      260 CCTCTCCT 267
```

RESULT 23
CB424734/c

LOCUS CB424734 294 bp mRNA linear EST 25-MAR-2003

DEFINITION 599034 MARC 6BOV Bos taurus cDNA 3', mRNA sequence.
 ACCESSION CB424734
 VERSION CB424734.1 GI:29195073
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 294)
 AUTHORS Smith,T.P.L., Roberts,A.J., Echternkamp,S.E., Chitko-McKown,C.G.,
 Wray,J.E. and Keele,J.W.
 TITLE A second set of bovine ESTs from pooled-tissue normalized libraries
 JOURNAL Unpublished (2003)
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_alt option. Vector identified with
 cross_match v0.990329.
 Plate: FQY8007 row: I column: 15
 Seq primer: TAGAAGGCACAGTCGAGG.

FEATURES Location/Qualifiers
 source 1. .294
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 6BOV"
 /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
 Library made with RNA pooled from multiple tissues
 including liver, lung, hypothalamus, pituitary, and
 placenta/endometrium."

ORIGIN

Query Match 2.7%; Score 43; DB 14; Length 294;
 Best Local Similarity 49.8%; Pred. No. 0.35;
 Matches 106; Conservative 1; Mismatches 106; Indels 0; Gaps 0;

```

Qy      1343 TGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTAAGTTGTAGAGAGGCAGCCATGC 1402
          ||||| :| | | | | |||| | | | | | | |||| |
Db      268 TGGCTGTGATTTAACAAAATGATTAAAGTGTTACCTACATGTGTAGCCGAAGTAGTGTGC 209

Qy      1403 ATTTGGCATTGGAATACAATCTGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACCAAA 1462
          | | | | | | | | | | | | | | | | | | | | |
Db      208 AGTGAGGTGTTTCTGAATACATGGTCAGATTTTGGAAAAAACAAAAACAAAAAAACA 149

Qy      1463 GTGAAATCTTGAGGAAAAATCCCTGGAAAGAGTGGAAGTCCTGCCTAACACGTAAGTGCC 1522
          ||| | | | | |||| | | | | | |||| | | | | |
Db      148 AGTAAAGTTCAACAACCATCCAACGAGAAAATTGCAAGTAGTGTGACAGAGCTGATTGAT 89

Qy      1523 TTCTTTGCTTGTTTGATTGACTGTGATGCTAGA 1555
          || ||||| | |||| | | | | |
  
```

Db 88 TTTGTTGCTTTCTTGATTTTTTTTTTTTCAGA 56

RESULT 24

CNS005TE

LOCUS	CNS005TE	997 bp	DNA	linear	GSS 03-JUN-1999
-------	----------	--------	-----	--------	-----------------

DEFINITION *Drosophila melanogaster* genome survey sequence TET3 end of BAC # BACR12K22 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.

ACCESSION AL060767

VERSION AL060767.1 GI:4943573

KEYWORDS GSS.

SOURCE *Drosophila melanogaster* (fruit fly)

ORGANISM *Drosophila melanogaster*

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 997)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

Location/Qualifiers

1. .997

```
/organism="Drosophila melanogaster"
```

```
/mol type="genomic DNA"
```

```
/db xref="taxon:7227"
```

```
/clone="BACR12K22"
```

```
/clone lib="RPCI-98"
```

```
/note="end : TET3"
```

ORIGIN

Query Match 2.7%; Score 42.4; DB 29; Length 997;

Best Local Similarity 25.2%; Pred. No. 1.1;

Matches 53; Conservative 70; Mismatches 87; Indels 0; Gaps 0;

Qy 22 CCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGCCATGACCAGTGCTGTTTGTGC 81
 |:| :|| | : | : | : ||| :|| :|:| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

Db 772 CYCYYYCCYYYCYTCYTYYYYYCTYYYYTYTYTTTYTYCCYTYTYTCTYYYTYTYYYCY 831

Qy 82 CCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTCTTTGCTCCTTAGAGCTGGGGC 141

```

      |: :: |      |: ::| | |: ::      ::| |: : |: |:| :: :|:      :
Db      832 CYCYCTYCCCTCYTYYYYCTYCYYYYYCYYYYYTCYTYTMTYTYYYYTYTYTYTYTHYTT 891

Qy      142 ACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCTYTCTGGCAAACACTTCCT 201
      |      ||: :| ||:: :| | ::      :::: : || :||| | :      |: ||
Db      892 YTTTTYCCYCYCCCTSYCYCTYCTYTYYYYTYYYTTTYTYTCTCTYYTCTYTTYTCT 951

Qy      202 ATAAACACACCGTGTGTCTGCCTATTGTC 231
      |      : : : | | :||: |: :| | ::
Db      952 YTCYTYYYYTYTYTYTCYTCYCYCYYYTCYY 981

```

RESULT 25

AA543856/c

LOCUS AA543856 629 bp mRNA linear EST 01-AUG-1997

DEFINITION vk34a07.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:948468 5', mRNA sequence.

ACCESSION AA543856

VERSION AA543856.1 GI:2292333

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 629)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:545324

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 459.

FEATURES

source

Location/Qualifiers

1. .629

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:948468"

/sex="male"

/tissue_type="mammary gland"

/dev_stage="4 weeks"

/lab_host="DH10B"

/clone_lib="Soares_mammary_gland_NbMMG"

/note="Organ: mammary gland; Vector: pT7T3D-Pac"

(Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCAATGGTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

ORIGIN

```
Query Match      2.6%; Score 40.6; DB 9; Length 629;
Best Local Similarity 55.2%; Pred. No. 2.9;
Matches    79; Conservative    0; Mismatches   64; Indels       0; Gaps        0;
```

Qy 408 TGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCT 467
| | | | ||||| | | | | | | | | | |

Db 580 TCCATTGTCTAGCCAGGAAGTGCTTAGCCTTAGACAGACACAGTGGAGTCTGAGTCACA 521

Qy 468 CACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCT 527
| | | | | | | | | | | | | | | | | |

Db 520 CAGTCCATCTCAGCCTCTCTGAGCTTCTGAGACATGGATCGAGACAGGGTACGGCGCAG 461

Qy 528 CGGCACAGCTTAGGTGTCCTGCA 550
| | | | | | | | | |

Db 460 GGGCCCCGGGTTTGCTGACTGGCA 438

RESULT 26

```

CB417759
LOCUS           CB417759                      294 bp      mRNA      linear      EST 25-MAR-2003
DEFINITION     590490 MARC 6BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      CB417759
VERSION        CB417759.1  GI:29181135
KEYWORDS       EST.
SOURCE         Bos taurus (cow)
  ORGANISM     Bos taurus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
                Bovidae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 294)
  AUTHORS      Smith,T.P.L., Roberts,A.J., Echternkamp,S.E., Chitko-McKown,C.G.,
                Wray,J.E. and Keele,J.W.
  TITLE        A second set of bovine ESTs from pooled-tissue normalized libraries
  JOURNAL       Unpublished (2003)
COMMENT        Contact: Smith TPL
                USDA, ARS, US Meat Animal Research Center
                PO Box 166, Clay Center, NE 68933-0166, USA
                Tel: 402 762 4366
                Fax: 402 762 4390
                Email: smith@email.marc.usda.gov
                Single pass sequencing. Bases called with phred v0.020425.c and
                trimmed with the aid of the trim_alt option. Vector identified with
                cross_match v0.990329.
                Plate: FQY8007 row: I column: 15
                Seq primer: GTAATACGACTCACTATAGGG.

```

```

FEATURES             Location/Qualifiers
     source            1. .294
                        /organism="Bos taurus"
                        /mol_type="mRNA"
                        /db_xref="taxon:9913"
                        /tissue_type="pooled"
                        /lab_host="DH10B"
                        /clone_lib="MARC 6BOV"
                        /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
                        Library made with RNA pooled from multiple tissues
                        including liver, lung, hypothalamus, pituitary, and
                        placenta/endometrium."

```

```

Query Match          2.5%; Score 39.8; DB 14; Length 294;
Best Local Similarity 48.8%; Pred. No. 3.2;
Matches 104; Conservative 1; Mismatches 108; Indels 0; Gaps 0;

Qy      1343 TGGCTCYGTCTAAGCACAAATGTTTAAGAAAGTRAGTTTAAGTTGTAGAGAGGCAGCCATGC 1402
          ||| | :| | | || ||| ||| | | | |||| | |||
Db      27 TGGGTGTGATTTAACAAAATGATTAAAGTGTTACCTACATGTGTAGCCGAAGTAGTGTGC 86

Qy      1403 ATTTGGCATTTTGAATACAATCTGGTGACTTGCTCTGGCTGCCAATAGAACCTAGTACCAAA 1462
          || | | || | | |||| | || | ||| || | || | | | | | |
Db      87 AGTGAGGTGTTTCTGAATACATGGTCAGATTTTTTGGAAAAAAACAAAAACAAAAAAACA 146

Qy      1463 GTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAGTCCTGCCTAACACGTAAGTGCC 1522
          ||| | | | |||| || | | | | ||| || | | | | |||
Db      147 AGTAAAGTTCAACAACCATCCAACGAGAAAATTGCAAGGAGTGTGACAGAGCTGATTGAT 206

Qy      1523 TTCTTTGCTTGTTTGATTGACTGTGATGCTAGA 1555
          || ||||| ||||| | | | |||
Db      207 TTTGTTGCTTTCTTGATTTTTTTTTTTTTTCAGA 239

```

```

LOCUS      AQ977239                      455 bp      DNA      linear      GSS 29-JAN-2000
DEFINITION RPCI-23-319D17.TJ RPCI-23 Mus musculus genomic clone
            RPCI-23-319D17, genomic survey sequence.
ACCESSION  AQ977239
VERSION    AQ977239.1  GI:6809540
KEYWORDS   GSS.
SOURCE     Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1  (bases 1 to 455)
  AUTHORS  Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
            Akinret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE    Mouse BAC End Sequences from Library RPCI-23
  JOURNAL  Unpublished (1999)
COMMENT    Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA

```


KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 662)
 AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
 Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
 Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
 Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
 Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
 Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
 TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. . 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
 Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
 and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
 Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
 Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A.,
 Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K.,
 Arakawa,T., Ishii,Y. and Hayashizaki,Y.
 Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
 Func. Genomics 2 pre, L72-L86 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.
 FEATURES Location/Qualifiers
 source 1. .662
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4632410H18"
 /sex="mixed"

/tissue_type="skin"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, 0 day neonate skin"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

ORIGIN

Query Match 2.5%; Score 39.4; DB 10; Length 662;
 Best Local Similarity 51.4%; Pred. No. 7;
 Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy	1388	GAGAGGCAGCCATGCATTTGGCATTTGAATACAATCTGGTGACTTGTCTGGCTGCCAATA	1447
Db	466	GAGTGGCACCCATGGTTCTTGGGGTCTCTAAATGAATGCTTAATTCTTCTTAATTAAAAAG	407
Qy	1448	GAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAGTCCTGCC	1507
Db	406	GAAGCTGGAGGAAGAGTAGTGTAGTGAGGAGAAGCCTAGAGAAGGGTTCTGTGTGCTGAC	347
Qy	1508	TAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCAAACC	1564
Db	346	TGACTAATGGCTGCCATTTATTTTGGTATTGTTTATGGCAGGCGAGGGACCAGAGC	290

RESULT 30

BX414498

LOCUS BX414498 1141 bp mRNA linear EST 15-MAY-2003

DEFINITION BX414498 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YI15
 5-PRIME, mRNA sequence.

ACCESSION BX414498

VERSION BX414498.1 GI:30769188

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1141)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 1974.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0CAP001AE08QP1&cluster=1974.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0CAP001AE08QP1.

FEATURES Location/Qualifiers
 source 1. .1141
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0CAP001YI15"
 /tissue_type="THYMUS"
 /clone_lib="Homo sapiens THYMUS"
 /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 2.5%; Score 39.4; DB 13; Length 1141;
 Best Local Similarity 34.1%; Pred. No. 9.8;
 Matches 98; Conservative 46; Mismatches 143; Indels 0; Gaps 0;

```

Qy      1193 AATCATGCCAGCAGAAAGTGGGACAGGCCAAATCCTCAAAGATGTCTCCTTGTACATCGAGA 1252
          || | || ||:| : ||:| || | : || |: ||| | : :: |
Db      26  AAAAAAGCAGGCWGGWACCGGWCCGGAATWCCCGGGAWATCGTCGACSSASGSGDSSGGS 85

Qy      1253 GTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCC 1312
          | :|| | | | | || | |: | | || | ||:|: : | | |
Db      86  GSGCGGCAGGAAGGGACGGCAGTCDCGCGCGGKGAGGAGCCGGGGKGGGGAGCGGCKC 145

Qy      1313 TGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAG 1372
          : | :| :||| ||| |:| |:| : : | | || || | : : ||
Db      146 GKGGAGGCKACKGCAGCACKGGGGKGKCGATKGTGKGKCCGACCCAGAACGCKKCKAGKKC 205

Qy      1373 TRAGTTTAAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTGTAATACAATCTGGTGACTT 1432
          :: : : | :|:|:| | | |: : :| | ::|:| | | : :|
Db      206 KGCKCKGCAAGGAKAKAKAAGAACKGAKKGGKGKGGCCGKKKAAKAAAAGAAKAGGAAA 265

Qy      1433 GTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAA 1479
          : | | | ||| : | | :||| : : | : ||
Db      266 CKGAACAGCCAGAAGAAACCKKCCCGAACACKGAAACSAKGGKGAA 312
  
```

RESULT 31

BI028780

LOCUS BI028780 361 bp mRNA linear EST 14-JUN-2001

DEFINITION CM0-MT0180-230201-789-g06 MT0180 Homo sapiens cDNA, mRNA sequence.

ACCESSION BI028780

VERSION BI028780.1 GI:14435410
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 361)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
 O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-MT0180-230201-789-g06&t3=2001-02-23&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 21
 High quality sequence stop: 73.

FEATURES Location/Qualifiers
 source 1. .361
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="MT0180"
 /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No.
 196,716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

ORIGIN

Query Match 2.5%; Score 39.2; DB 12; Length 361;
 Best Local Similarity 50.0%; Pred. No. 5.5;
 Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 439 TTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGG 498
 ||| ||| ||| | ||| |||| | | |||| | | ||| | ||
 Db 154 TTCGGAGAGCAGGGTGAGAGAGAGAGATATGGAGAAACAGTGCACCAGCGAGATGGATGA 213

Qy 499 AGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCT 558
 | | | | | | | | | | | | | | | | | | | | | |
 Db 214 GGGATGGGGGAGAGATGGGGACGGGGTGAGGGCACCCCTGGAGGGGGACGCACAGGGCCAG 273
 Qy 559 ACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTG 618
 | | | | | | | | | | | | | | | | | | | | | |
 Db 274 AGAGAGACAGGAGAGGCTGAACCAAGAGTCAAGCACACACATAGCTGTGTGTGGGTGCTG 333
 Qy 619 AATGGGTGGGTGGGCC 634
 | | | | | | | | | | | | | | | | | | | | | |
 Db 334 GATGGGTGCGGGGGAC 349

RESULT 32

CK203027/c

LOCUS CK203027 837 bp mRNA linear EST 08-DEC-2003

DEFINITION FGAS011553 Triticum aestivum FGAS: Library 3 Gate 6 Triticum
 aestivum cDNA, mRNA sequence.

ACCESSION CK203027

VERSION CK203027.1 GI:39565417

KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooideae; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 837)

AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
 Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,
 Penniket, C., Roach, J.L. and Sarhan, F.

TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops

JOURNAL Unpublished (2003)

COMMENT Contact: Wm L Crosby

Bioinformatics

University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada

Tel: 306 966 1769

Fax: 306 966 2033

Email: fgas_ests@cs.usask.ca

This sequence is the direct result of the Base calling software
 Phred (default parameters). It is the raw base calls. To aid in the
 identification of the high quality insert the software Lucy
 (default parameters) has been run on this sequence. Lucy identified
 the region [102,331].

Plate: L3C116 row: J column: 21.

FEATURES

source

Location/Qualifiers

1. .837

/organism="Triticum aestivum"

/mol_type="mRNA"

/db_xref="taxon:4565"

/clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"

/note="Organ: Root; Vector: pCMV.SPORT6; Root tissue from
 control, cold-acclimated and salt stressed wheat cultivar
 Norstar. 7 mRNA populations were combined before
 constructing the library; 7 day non-acclimated roots, 1,

23, and 53 days cold-acclimated at 4C, and 30 minutes, 3 hours and 6 hours treated roots with 200mM NaCl. Non-acclimated and cold-acclimated plants were grown in vermiculite while salt stressed plant were grown hydroponically. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

ORIGIN

Query Match 2.5%; Score 39.2; DB 14; Length 837;
 Best Local Similarity 51.6%; Pred. No. 9.3;
 Matches 80; Conservative 3; Mismatches 72; Indels 0; Gaps 0;

Qy 1231 GATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAG 1290
 || || || || | ||||| | | ||| | ||| ||
 Db 455 GAGGTGGCCGTGTTGCTGCTTGGTGGCCAGGGCAAATAAGAAGTAGGAGGGCCAGGGGAG 396

Qy 1291 TGCCTGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYG 1350
 ||||| | : : | | | | | ||| | ||||| | || |
 Db 395 GGCCTGGGGTGCACTGCCCCGGCGGAAGAAGTTCGGAGGATTTGGGAGGCAGAAGGTCGG 336

Qy 1351 TCTAAGCACAAATGTTTAAGAAGTRAGTTTAAGTTG 1385
 || || | || | | : |||| | | |
 Db 335 TCGAAGAGGGAGGTGTCCGTTGAAGTTTTGGTNG 301

RESULT 33

AA914287/c

LOCUS AA914287 458 bp mRNA linear EST 14-APR-1998
 DEFINITION vy99b08.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
 IMAGE:1314327 5' similar to TR:O15273 O15273 TELETHONIN. ;, mRNA
 sequence.
 ACCESSION AA914287
 VERSION AA914287.1 GI:3053679
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 458)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:684623

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 449.

FEATURES Location/Qualifiers
 source 1. .458
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1314327"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /clone_lib="Soares_mammary_gland_NbMMG"
 /note="Organ: mammary gland; Vector: pT7T3D-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer [5'
 TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT7T3 vector.
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
 constructed and normalized by Bento Soares and M.Fatima
 Bonaldo."

ORIGIN

Query Match 2.5%; Score 39; DB 9; Length 458;
Best Local Similarity 54.5%; Pred. No. 7.3;
Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy      408 TGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCT 467
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      218 TCCATTGTCCTAGCCAGGAAGTGCCTAGCCTGAGACAGACACAGTGGAGTCTGAGTCACA 159

Qy      468 CACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCT 527
          || | | | | | | | | | | | | | | | | | | | | | | | |
Db      158 CAGTCCATCTCAGCCTCTCTGAGCTTCCTGAGACATGGATCGAGACAGGGTACGGCGCAG 99

Qy      528 CGGCACAGCTTAGGTGTCCTGCA 550
          ||| | | | | | | | | |
Db      98 GGGCCCGGGTTTGCTGACTGGCA 76
```

RESULT 34

AA060852/c

LOCUS AA060852 493 bp mRNA linear EST 23-SEP-1996

DEFINITION mj86d02.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
 IMAGE:482979 5', mRNA sequence.

ACCESSION AA060852

VERSION AA060852.1 GI:1554690

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

 ORGANISM Mus musculus

 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 493)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:293723
Putative full length read
vector to vector length is 512
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 481.

FEATURES Location/Qualifiers

source 1. .493
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:482979"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3NMF19.5"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGCCGCATTTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

Db 55 GGGCCCCGGGTTTGCTGACTGGCA 33

RESULT 35

AA882149/c

LOCUS AA882149 525 bp mRNA linear EST 26-MAR-1998

DEFINITION vx38e02.rl Stratagene mouse lung 937302 Mus musculus cDNA clone
IMAGE:1277498 5' similar to TR:015273 O15273 TELETHONIN. ;, mRNA
sequence.

ACCESSION AA882149

VERSION AA882149.1 GI:2991260

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 525)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:669298

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 493.

FEATURES

source

Location/Qualifiers

1. .525

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6 x CBA"

/db_xref="taxon:10090"

/clone="IMAGE:1277498"

/sex="female"

/tissue_type="lung"

/dev_stage="6-8 month old"

/lab_host="SOLR (kanamycin resistant)"

/clone_lib="Stratagene mouse lung 937302"

/note="Organ: lung; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. 6-8 month old female lung and 1.5 year old male

lung were source of mRNA. Average insert size: 1.5 kb;

Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG

3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match

2.5%; Score 39; DB 9; Length 525;

Best Local Similarity 54.5%; Pred. No. 7.9;
Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

Qy      408 TGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCT 467
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      263 TCCATTGTCCCTAGCCAGGAAGTGCCTAGCCTGAGACAGACACAGTGGAGTCTGAGTCACA 204

Qy      468 CACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCT 527
          || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      203 CAGTCCATCTCAGCCTCTCTGAGCTTCCTGAGACATGGATCGAGACAGGGTACGGCGCAG 144

Qy      528 CGGCACAGCTTAGGTGTCCTGCA 550
          ||| | | | | | | | | | | | | | | |
Db      143 GGGCCCGGGTTTGCTGACTGGCA 121

```

RESULT 36

BI416074/c

LOCUS BI416074 856 bp mRNA linear EST 14-AUG-2001

DEFINITION 602987346F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5143451 5', mRNA sequence.

ACCESSION BI416074

VERSION BI416074.1 GI:15176984

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 856)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11353 row: 1 column: 12

High quality sequence start: 17

High quality sequence stop: 856.

FEATURES

source

Location/Qualifiers

1. .856

/organism="Mus musculus"

/mol_type="mRNA"

/strain="Czech II"

/db_xref="taxon:10090"

/clone="IMAGE:5143451"

/tissue_type="pooled lung tumors"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NCI_CGAP_Lu33"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st


```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30287594"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
Normalized, full-length enriched library from pool of
mouse embronic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTTCTAGATCGCGAGCGGCCGCC(T)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5 . Tissue contributed by
David Rowe; library constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

```

ORIGIN

```

Query Match          2.5%;  Score 39;  DB 14;  Length 929;
Best Local Similarity 54.5%;  Pred. No. 11;
Matches 78;  Conservative 0;  Mismatches 65;  Indels 0;  Gaps 0;

Qy      408 TGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCT 467
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      597 TCCATTGTCCTAGCCAGGAAGTGCCTAACCTGAGACAGACACAGTGGAGTCTGAGTCACA 538

Qy      468 CACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCT 527
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      537 CAGTCCATCTCAGCCTCTCTGAGCTTCCTGAGACATGGATCGAGACAGGGTACGGCGCAG 478

Qy      528 CGGCACAGCTTAGGTGTCCTGCA 550
      | | | | | | | | | | | | | |
Db      477 GGGCCCGGGTTTGCTGACTGGCA 455

```

RESULT 38 AK010167/c

```

LOCUS      AK010167          933 bp    mRNA    linear    HTC 20-SEP-2003
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310075E03 product:titin-cap, full insert sequence.
ACCESSION  AK010167
VERSION    AK010167.1  GI:12845417
KEYWORDS   HTC; CAP trapper.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  1
AUTHORS    Carninci,P. and Hayashizaki,Y.
TITLE      High-efficiency full-length cDNA cloning
JOURNAL    Meth. Enzymol. 303, 19-44 (1999)
MEDLINE    99279253
PUBMED     10349636
REFERENCE  2
AUTHORS    Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

```

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 933)
 AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
 Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
 Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
 Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
 Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
 Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
 Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
 Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
 Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
 Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission
 JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer

[5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```

FEATURES                      Location/Qualifiers
    source                      1. .933
                                /organism="Mus musculus"
                                /mol_type="mRNA"
                                /strain="C57BL/6J"
                                /db_xref="FANTOM_DB:2310075E03"
                                /db_xref="MGI:1910392"
                                /db_xref="taxon:10090"
                                /clone="2310075E03"
                                /sex="male"
                                /tissue_type="tongue"
                                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                                /dev_stage="adult"
    CDS                          25. .528
                                /note="unnamed protein product; putative
                                titin-cap (MGD|MGI:1330233)"
                                /codon_start=1
                                /protein_id="BAB26743.1"
                                /db_xref="GI:12845418"
                                /translation="MATSELSCQVSEENQERREAFWAEWKDLTLSTRPEEGCSLHEED
                                TORHETYHRQGQCQAVVQVQSPWLVMRLGILGRGLQEYQLPYQRVLPPIFTPTKVGAS
                                KEEREETPIQLRELLALETALGGQCVERQDVAEITKQLPPVVPVSKPGPLRRTLRSRM
                                SQEAQRG"

```

ORIGIN

Query Match 2.5%; Score 39; DB 11; Length 933;
 Best Local Similarity 54.5%; Pred. No. 11;
 Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

Qy      408 TGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCT 467
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      597 TCCATTGTCCTAGCCAGGAAGTGCCTAGCCTGAGACAGACACAGTGGAGTCTGAGTCACA 538

Qy      468 CACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCT 527
          || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      537 CAGTCCATCTCAGCCTCTCTGAGCTTCCTGAGACATGGATCGAGACAGGGTACGGCGCAG 478

Qy      528 CGGCACAGCTTAGGTGTCTCTGCA 550
          || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      477 GGGCCCGGGTTTGCTGACTGGCA 455

```

RESULT 39
 BC032992/c

LOCUS BC032992 946 bp mRNA linear HTC 20-SEP-2002
 DEFINITION Mus musculus, clone IMAGE:1281423, mRNA.
 ACCESSION BC032992
 VERSION BC032992.1 GI:21426937
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)

ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 946)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (14-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: Marcello Bento Soares, Ph.D. cDNA Library Preparation: Soares Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology http://www.systemsbiology.org contact: amadan@systemsbiology.org Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 66 Row: k Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: no 5' EST match.

FEATURES	Location/Qualifiers
source	1..946 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:1281423" /tissue_type="Thymus gland, mouse" /clone_lib="Soares_thymus_2NbMT" /lab_host="DH10B" /note="Vector: pT7T3D-Pac"

ORIGIN

Query Match 2.5%; Score 39; DB 11; Length 946;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy		408 TGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCT	467
Db		581 TCCATTGTCCTAGCCAGGAAGTGCTTAGCCTGAGACAGACACAGTGGAGTCTGAGTCACA	522
Qy		468 CACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCT	527
Db		521 CAGTCCATCTCAGCCTCTCTGAGCTTCCTGAGACATGGATCGAGACAGGGTACGGCGCAG	462
Qy		528 CGGCACAGCTTAGGTGTCCTGCA	550
Db		461 GGGCCCCGGGTTTGCTGACTGGCA	439

RESULT 40
 CNS0037Q/c
 LOCUS CNS0037Q 1101 bp DNA linear GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K14 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL064465
 VERSION AL064465.1 GI:4941722
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES
 Location/Qualifiers
 source 1. .1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR08K14"
 /clone_lib="RPCI-98"
 /note="end : TET3"
 ORIGIN
 Query Match 2.5%; Score 39; DB 29; Length 1101;
 Best Local Similarity 12.9%; Pred. No. 13;
 Matches 59; Conservative 188; Mismatches 210; Indels 0; Gaps 0;
 Qy 409 GCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTC 468
 ::: :: :::: : : : :|: : : : : : : : :|::: :
 Db 1034 KMKNBMKNBVKVKMKCKBABNKCKMKMNCKMBMKNVBGBKCBNMMCKASCMGSBMSCS 975
 Qy 469 ACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTC 528
 :: : ::||::: : : | : : : : : : : : :| : : :

Db 974 CSRCKCKNKKKBKBTCKBKKBKKBKBTBTMBMBKNBKYKBKKYNKCNKMCBYDCBBCY 915
 Qy 529 GGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGACCTCCACAGCA 588
 :| : : : : : | : : | : : : :| : : : | : : :
 Db 914 CKCKHKYKCKCKVKVBDAADAKNKNANAAAAAAMDHMDVMBAMBSNBMANCKCMSMK 855
 Qy 589 AAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGT 648
 : : : :|| | : : : : : :| : : : : : :| : : :
 Db 854 CNCKNBVBKNBANDCTCNTKWYTWDYKYYKTHTMKBYKTCYMTMBYYTCTWYATMKTY 795
 Qy 649 TGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTC 708
 | || : :| : : || : : : : : :| :| : :| :|| : || : |
 Db 794 TMTCBTCTYAKTWTATMTCHKMKHMMMDMWCKMKCKMHMATMACMMMNMTMTYTMTT 735
 Qy 709 CTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAACATGCCCTGCCTGAAGCCGCTT 768
 :| : | :| | : : : : : : : : : : :| : : : : :
 Db 734 YTYKTAYTKTTCTYTKBTKYAMAKAHAATTMBNHVWBYCWMTMTMNNYCAHMMMKTDW 675
 Qy 769 GCTGCTTCTCACTGATTTCGCTCTCCCTTCCTTGACTCGCCACCACCTGTCTGTGT 828
 : : :| :| : : | :| : :| : :| : : :| :| : : : :
 Db 674 AMKTWCNTGTAYKYAKNHTTCNTBTSTWKMNCMYBHMYCHMNTTRYMTCCCHCTCAYKYAH 615
 Qy 829 AGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGC 865
 : : : :|| :| : : : :| :| : : | ||||
 Db 614 RTSHRYDYTAGMADCTVDDRNRTRGVGDRRVGAGGGC 578

RESULT 41

BQ752298

LOCUS BQ752298 834 bp mRNA linear EST 18-JUL-2002
 DEFINITION EST632861 DSCT Colletotrichum trifolii cDNA clone pDSCT11-59, mRNA sequence.

ACCESSION BQ752298

VERSION BQ752298.1 GI:21907703

KEYWORDS EST.

SOURCE Colletotrichum trifolii

ORGANISM Colletotrichum trifolii

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
 mitosporic Phyllachoraceae; Colletotrichum.

REFERENCE 1 (bases 1 to 834)

AUTHORS Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
 Cheung,F. and Fraser,C.M.

TITLE ESTs from mycelia of Colletotrichum trifolii race 1

JOURNAL Unpublished (2002)

COMMENT Other ESTs: EST632862

Contact: Deborah A. Samac

Department of Plant Pathology

University of Minnesota

495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA

Tel: 612 625 1243

Fax: 651 649 5058

Email: debbys@puccini.crl.umn.edu

TIGR sequence name: MTSAK59TK More information is available at:

www.medicago.org

Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers

source 1. .834
 /organism="Colletotrichum trifolii"
 /mol_type="mRNA"
 /strain="race 1"
 /db_xref="taxon:5466"
 /clone="pDSCT11-59"
 /tissue_type="mycelia"
 /dev_stage="Young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium containing 2%glucose)."
 /lab_host="DH5alpha"
 /clone_lib="DSCT"
 /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+ enriched RNA The cDNA was ligated into Lambda gtl1 from Stratagene and packaged using Gigapack packaging extracts. An aliquot of the amplified library was used to transduce E. coli Y1090 and phage DNA was purified from a liquid lysate. The cDNA inserts were gel purified after EcoRI digestion and ligated into pBluescript SK+. Aliquots of the ligation were used to transform E. coli DH5alpha which were plated onto medium with X-gal for selection of recombinants."

ORIGIN

Query Match 2.5%; Score 38.8; DB 13; Length 834;
 Best Local Similarity 53.2%; Pred. No. 12;
 Matches 82; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 431 GCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAG 490
 |||| | | | | | | | | | | | | | | | |
 Db 424 GCTGGGAGGTCGCGGCCCGGGTGGCGCGCCGAGGCTTTCCATCGACGGCGGCTTGTTGAG 483

Qy 491 CTCCCTGGAGCAAGGTTTCGGTTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCA 550
 |||| | | | | | | | | | | | | | | | |
 Db 484 CTCCGTTCCGGAAGGTCCGCCCCGAGGCGCGCTCGCTCGTCCCATAGGCGGCGTCATTCA 543

Qy 551 TGTGTCCTACAGCGTCAGGTAAGGGGACCTCCAC 584
 || || | || | | | | | | | | | |
 Db 544 TGC GTGGGGCGGCGGCGGCGTTCCCGTCGTCGAC 577

RESULT 42 AA390068/c

LOCUS AA390068 551 bp mRNA linear EST 23-APR-1997
 DEFINITION mv35b05.r1 GuayWoodford Beier mouse kidney day 0 Mus musculus cDNA clone IMAGE:657009 5' similar to gb:J05021 EZRIN (HUMAN); gb:X60671 M.musculus mRNA for ezrin (MOUSE);, mRNA sequence.

ACCESSION AA390068
 VERSION AA390068.1 GI:2043083
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 551)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:402857

High quality sequence stop: 279.
FEATURES Location/Qualifiers
source 1. .551
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:657009"
/tissue_type="kidney"
/dev_stage="newborn (day 0)"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="GuayWoodford Beier mouse kidney day 0"
/note="Organ: kidney; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3' Library provided
Lisa Guay-Woodford."

ORIGIN

Query Match 2.5%; Score 38.6; DB 9; Length 551;
Best Local Similarity 53.0%; Pred. No. 11;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 27 TCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGCCATGACCAGTGCTGTTTGTGCCCTTT 86
|| ||| || ||| ||||| ||| | ||| ||| ||| |||
Db 197 TCACCAGGTGCAGCTCCTCTTTGGTCTTCACCAGGTCGTCCTGGGCTTCTTTAGCCCCGT 138
QY 87 GTGTGGCCTCCCCTGCTGTTGGGCTCTCTGTCTTTGCTCCTTAGAGCTGGGGCACCTG 146
| ||| || | || || || || || ||| || | ||
Db 137 GCTGCCACTCTTCTACCTCGTCCTCCNNTCGCTCCGCGCCTCCTCCAGCAGTGCGATCTT 78
QY 147 AGCCCTCCTCTGTGCCAGCCTTTCTCCCAGC 177
||| | || ||||| | | |||||
Db 77 GGCCGTGTACTCTGCCAGCTCTGCAGCCAGC 47

RESULT 43

CNS00418

LOCUS CNS00418 987 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #

Qy 219 TCTGCCTATTGTC 231
 | | :| :: ::
 Db 910 TTTYTYTTYYYYY 922

RESULT 44

BX335650/c

LOCUS BX335650 1201 bp mRNA linear EST 02-MAY-2003

DEFINITION BX335650 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI017YH11 5-PRIME, mRNA sequence.

ACCESSION BX335650

VERSION BX335650.1 GI:30343426

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1201)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ InVitroGen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DI017CD06QP1.

FEATURES

source

Location/Qualifiers

1. .1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DI017YH11"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 2.5%; Score 38.6; DB 13; Length 1201;

Best Local Similarity 43.6%; Pred. No. 18;

Matches 78; Conservative 18; Mismatches 83; Indels 0; Gaps 0;

Qy 22 CCCATTCCACAGCTGGGTCTCTCTTTGGTTTTCTCAGCCATGACCAGTGCTGTTTGTGC 81
 |:| ::|| || | | || || ||| |||:| : : | || | ||| |
 Db 835 CYCTKKCCCTTCCTTTTTTTTTTTTTTTTTTTTTTYCCYCCYTTCCCCCCCCCTTTTTTTT 776
 Qy 82 CCTTTGTGTGGCCTCCCGTGCTGTTGGGCTCTCTCTGTCTTTGCTCCTTAGAGCTGGGGC 141
 :||| | | ||| | | || | | || | ||| |::: | ::|::|
 Db 775 TSTTTTTTTTYCTTCCCTTCTTTTTTTCTTTTTTCCCTTTTTTGCCYKKGCTKKKGKKC 716
 Qy 142 ACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCAGCATTCCTYTCTGGCAAACACTTCC 200
 :| ||| || || |: | || | ||| || :|:| | | | |||

RESULT 45

AA525033

LOCUS AA525033 407 bp mRNA linear EST 05-AUG-1997

DEFINITION nh36c06.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:954442, mRNA sequence.

ACCESSION AA525033

VERSION AA525033.1 GI:2265961

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 407)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 507 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

source

Location/Qualifiers

1. .407

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:954442"

/sex="Male"

/dev_stage="45 years old"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Pr3"

/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected cells

histologically-determined to be fully malignant prostate cancer cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

ORIGIN

Query Match

2.4%; Score 38.4; DB 9; Length 407;

Best Local Similarity 49.5%; Pred. No. 10;
Matches 99; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```

Qy      741 TTTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTC 800
        |||| | | | || || | | | | | | | | | | | | | |
Db      108 TTAATATACTCTATGGATGACCCAGCAAGTTTGCTGTTTCAGAATCCTCCTCTTCTGTTT 167

Qy      801 CTTGACTCGCCACACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTG 860
        |||| | | || || | | || | | |||| | | ||
Db      168 TTTGAACTTTGAAAACAAAAGATGTGCTGGGAGACGCGGCCCTAGAGTGTGCTTACTC 227

Qy      861 GGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGC 920
        || | | || | || || | |||| | | | | |||| | |
Db      228 CAGGTCCTTGATTGTCCAGACTGTGGAGGGGGAAGGGCAGATCTATGCCAAGAGGGGAAC 287

Qy      921 AGTGTGCAGGTGTACCATCT 940
        || || || | || ||
Db      288 AGGCTGTAGAGGCCACAGCT 307

```

RESULT 46

AA524916

LOCUS AA524916 412 bp mRNA linear EST 05-AUG-1997

DEFINITION nh31a09.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:953944, mRNA sequence.

ACCESSION AA524916

VERSION AA524916.1 GI:2265844

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 412)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 501 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 340.

FEATURES

source

Location/Qualifiers

1. .412

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:953944"

/sex="Male"

```

/dev_stage="45 years old"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr3"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected cells
histologically-determined to be fully malignant prostate
cancer cells. Double-stranded cDNA was ligated to EcoRI
adaptors, 5 cycles of PCR applied to the cDNA with an
adaptor-specific primer, and the resulting PCR product
subcloned into pAMP10 by the UDG-cloning method (Life
Technologies). Average insert size is 600 bp. NOTE: Not
directionally cloned. This library was constructed by
David Krizman."

```

ORIGIN

```

Query Match          2.4%; Score 38.4; DB 9; Length 412;
Best Local Similarity 49.5%; Pred. No. 10;
Matches 99; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```

```

Qy      741 TTAAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTC 800
        |||| | | | || ||| | || | | | | | | | | | |
Db      108 TTAATATACTCTATGGATGACCCAGCAAGTTTGCTGTTTCAGAATCCTCCTCTTCTGTTT 167

Qy      801 CTTGACTCGCCCAACACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTG 860
        |||| | | || |||| | || ||| | |||| | | ||
Db      168 TTTGAACTTTTCGAAAACAAAAGATGTGCTGGGAGACGCGGCCCTAGAGTGTGCTTACTC 227

Qy      861 GGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGC 920
        || | || | || || || | |||| | | | | |||| | |
Db      228 CAGGTCTTGATTGTCCAGACTGTGGAGGGGGAAGGGCAGATCTATGCCAAGAGGGGAAC 287

Qy      921 AGTGTGCAGGTGTACCATCT 940
        || || || | || ||
Db      288 AGGCTGTAGAGGCCACAGCT 307

```

RESULT 47

CF486702

LOCUS CF486702 472 bp mRNA linear EST 08-SEP-2003

DEFINITION POL1_39_F01.b1_A002 Pollen Sorghum bicolor cDNA clone

POL1_39_F01_A002 3', mRNA sequence.

ACCESSION CF486702

VERSION CF486702.1 GI:34515571

KEYWORDS EST.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 472)

AUTHORS Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,
Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,
Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and
Pratt,L.H.

TITLE EST database from Sorghum: pollen

Db 308 GCTATGGCGTGCAGGAAAAAGGAGGGGTGCATGAAGCGTTTACGGAAGGGGTGAAGCGTT 367

Qy 1070 CAGG 1073

|||

Db 368 TAGG 371

RESULT 48

CNS0060N

LOCUS CNS0060N 910 bp DNA linear GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL065629

VERSION AL065629.1 GI:4944698

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 910)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers

source

1. .910
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14J21"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 2.4%; Score 38.4; DB 29; Length 910;
Best Local Similarity 22.7%; Pred. No. 17;
Matches 34; Conservative 68; Mismatches 48; Indels 0; Gaps 0;

Qy 41 TCTTCTTTGGTTTTCTCAGCCATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCT 100

/cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
 Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 2.4%; Score 38.2; DB 28; Length 472;
 Best Local Similarity 56.9%; Pred. No. 13;
 Matches 70; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 608 TGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAG 667
 || || ||| ||| ||| || | || | |||| ||| || ||
 Db 413 TGGTTTCTCTGGTTGGCGGGGGAGGGGGTGAATTTCTGCAGTTCTGTACAGGGGAGGAG 354

Qy 668 GGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGG 727
 || |||| | | |||| |||| ||| | | || |||| ||||
 Db 353 GGAGAACTAGGACCTCTGAAACAGGAGTGTGGCTGTGCCCCCTGGAGTCTGCAAAGGG 294

Qy 728 CAG 730
 ||
 Db 293 GAG 291

RESULT 50

BZ849786/c

LOCUS BZ849786 560 bp DNA linear GSS 18-MAR-2003

DEFINITION CH240_280D12.TV CHORI-240 Bos taurus genomic clone CH240_280D12,
 genomic survey sequence.

ACCESSION BZ849786

VERSION BZ849786.1 GI:29077187

KEYWORDS GSS.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 560)

AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
 Shvartsbeyn,A., Gebregeorgis,E., Chen,D., Riggs,F., de Jong,P.,
 Crawford,A.M. and McEwan,J.C.

TITLE Bovine BAC End Sequences from Library CHORI-240

JOURNAL Unpublished (2003)

COMMENT Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the bovine BAC library CHORI-240

(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/ordering_information.htm). This work
 was undertaken as part of the International Bovine BAC Mapping
 Consortium (IBBMC) by AgResearch Ltd., New Zealand and The

Institute of Genomic Research (TIGR), USA.

Plate: 280 row: D column: 12

Seq primer: T7

Class: BAC ends.

FEATURES
source Location/Qualifiers
1. .560
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_280D12"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 2.4%; Score 38.2; DB 28; Length 560;
Best Local Similarity 56.9%; Pred. No. 14;
Matches 70; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 608 TGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAG 667
|| || ||| ||| ||| || | || | |||| ||| || ||
Db 436 TGGGTTCTCTGGTTGGCGGGGGAGGGGGTGGAAATTCTGCAGTTCTGTACAGGGGAGGAG 377

Qy 668 GGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCTCCTCCTCCTCCTGCGCAGGG 727
|| |||| | | |||| |||| |||| | | | | || |||| ||||
Db 376 GGAGAAACTAGGACCTCTGAAACAGGAGTGTGGCTGTGCCCCCTGGAGTCCTGCAAAGGG 317

Qy 728 CAG 730
||
Db 316 GAG 314

Search completed: April 29, 2004, 18:39:37

Job time : 5121.38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 14:53:09 ; Search time 6697.85 Seconds
(without alignments)
10159.758 Million cell updates/sec

Title: US-09-989-981A-9_COPY_3436_5005
Perfect score: 1570
Sequence: 1 cgaagcatcctgaagtacag.....ctagagagcaaaccagagc 1570

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*

28: em_un:*
 29: em_vi:*
 30: em_htg_hum:*
 31: em_htg_inv:*
 32: em_htg_other:*
 33: em_htg_mus:*
 34: em_htg_pln:*
 35: em_htg_rod:*
 36: em_htg_mam:*
 37: em_htg_vrt:*
 38: em_sy:*
 39: em_htgo_hum:*
 40: em_htgo_mus:*
 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			%		Query		DB	ID	Description
	No.	Score	Match	Length					
	1	1568	99.9	6043	6	AX685737			AX685737 Sequence
	2	965.4	61.5	1000	10	F351786S01			AF351786 Mus muscu
c	3	868.6	55.3	237445	2	AC120701			AC120701 Rattus no
	4	868.6	55.3	312858	2	AC112747			AC112747 Rattus no
	5	868.2	55.3	40929	10	AY145899			AY145899 Rattus no
c	6	555.6	35.4	567	10	AF404108			AF404108 Mus muscu
c	7	412.4	26.3	588	10	AF404109			AF404109 Rattus no
	8	402.6	25.6	463	10	F351786S02			AF351787 Mus muscu
c	9	398.4	25.4	1314	10	F351799S01			AF351799 Mus muscu
	10	358.6	22.8	359	6	AX685738			AX685738 Sequence
c	11	299.4	19.1	185045	2	AC146466			AC146466 Callithri
c	12	298.8	19.0	178016	2	AC146787			AC146787 Aotus nan
	13	284.4	18.1	2351	10	AY195873			AY195873 Mus muscu
	14	284.4	18.1	2354	6	AX456524			AX456524 Sequence
	15	284.4	18.1	2354	10	AF312713			AF312713 Mus muscu
	16	282.8	18.0	2351	10	AY195872			AY195872 Mus muscu
c	17	278.6	17.7	127066	9	AC084265			AC084265 Homo sapi
c	18	278.6	17.7	139342	9	AC108476			AC108476 Homo sapi
	19	275.4	17.5	159346	2	AC145533			AC145533 Lemur cat
c	20	261.8	16.7	207760	2	AC146286			AC146286 Callicebu
	21	244.8	15.6	4899	9	AF404106			AF404106 Homo sapi
	22	242.8	15.5	5459	6	AX456521			AX456521 Sequence
c	23	241.2	15.4	2809	9	F351812S01			AF351812 Homo sapi
	24	238.8	15.2	202533	2	AC146464			AC146464 Saimiri s
	25	215	13.7	2512	6	AX747300			AX747300 Sequence
	26	215	13.7	2512	9	AK091997			AK091997 Homo sapi
	27	191.4	12.2	2258	6	AX320881			AX320881 Sequence
c	28	179.8	11.5	581	9	AF404107			AF404107 Homo sapi
	29	174	11.1	68166	2	AC084712			AC084712 Homo sapi
	30	173.6	11.1	2470	10	AF312714			AF312714 Rattus no
c	31	164	10.4	2284	10	AY196216			AY196216 Mus muscu
c	32	164	10.4	2285	10	AY196215			AY196215 Mus muscu
c	33	164	10.4	3674	10	AF324495			AF324495 Mus muscu

	34	151.2	9.6	226	6	BD223287	BD223287 Toxicolog
	35	150.8	9.6	235	6	AR121818	AR121818 Sequence
	36	146.4	9.3	1915	6	AX456523	AX456523 Sequence
	37	146.4	9.3	1959	6	AX685729	AX685729 Sequence
c	38	145.4	9.3	68166	2	AC084712	AC084712 Homo sapi
	39	135.8	8.6	2035	6	AX456526	AX456526 Sequence
	40	107	6.8	2516	6	AX456520	AX456520 Sequence
	41	107	6.8	2740	9	AF312715	AF312715 Homo sapi
	42	101.6	6.5	249	6	AX320886	AX320886 Sequence
	43	101.6	6.5	2340	6	AX320883	AX320883 Sequence
	44	101.6	6.5	2340	6	AX685733	AX685733 Sequence
	45	101.6	6.5	2340	9	AF320293	AF320293 Homo sapi
	46	93	5.9	1920	6	AX456519	AX456519 Sequence
	47	90	5.7	122	6	AX320887	AX320887 Sequence
c	48	84	5.4	4829	10	AF351785	AF351785 Rattus no
c	49	67.6	4.3	135280	2	AC146282	AC146282 Takifugu
c	50	63	4.0	2019	6	AX685731	AX685731 Sequence

ALIGNMENTS

RESULT 1

AX685737

LOCUS AX685737 6043 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 9 from Patent WO02081691.

ACCESSION AX685737

VERSION AX685737.1 GI:29371746

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 9 17-OCT-2002;
Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)

FEATURES

source

Location/Qualifiers

1. .6043

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

/note="ABCG8 exon 2 (reverse strand) through ABCG5 exon 2
(forward strand)"

ORIGIN

Query Match 99.9%; Score 1568; DB 6; Length 6043;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTCTTTGGTTTTCTCAGC 60
|||||

Db 3436 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTCTTTGGTTTTCTCAGC 3495

Qy 61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120

Db	3496	 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC	3555
Qy	121	TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT	180
Db	3556	 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT	3615
Qy	181	CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG	240
Db	3616	 CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG	3675
Qy	241	AACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTAT	300
Db	3676	 AACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTAT	3735
Qy	301	CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT	360
Db	3736	 CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT	3795
Qy	361	AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAG	420
Db	3796	 AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAG	3855
Qy	421	CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG	480
Db	3856	 CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG	3915
Qy	481	GGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAG	540
Db	3916	 GGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAG	3975
Qy	541	GTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCT	600
Db	3976	 GTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCT	4035
Qy	601	CTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGC	660
Db	4036	 CTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGC	4095
Qy	661	AGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTG	720
Db	4096	 AGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTG	4155
Qy	721	CGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCAC	780
Db	4156	 CGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCAC	4215
Qy	781	TGATTCTGCTCTCCCCCTCCTTGACTCGCCACACCTGTCTGTGTAGATGGAGAAGG	840
Db	4216	 TGATTCTGCTCTCCCCCTCCTTGACTCGCCACACCTGTCTGTGTAGATGGAGAAGG	4275
Qy	841	CTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGG	900
Db	4276	 CTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGG	4335
Qy	901	GTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAA	960

Db	4336	GTTCACCTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAA	4395
Qy	961	TCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACCTTGTCATTATACCTCC	1020
Db	4396	TCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACCTTGTCATTATACCTCC	4455
Qy	1021	AAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACT	1080
Db	4456	AAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACT	4515
Qy	1081	GGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTG	1140
Db	4516	GGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTG	4575
Qy	1141	CCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGCCTTGGTGGAACATCAAATCATGC	1200
Db	4576	CCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGCCTTGGTGGAACATCAAATCATGC	4635
Qy	1201	CAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAG	1260
Db	4636	CAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAG	4695
Qy	1261	ATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCCTGTACTTC	1320
Db	4696	ATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCCTGTACTTC	4755
Qy	1321	TAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAATGTTTAAGAAGTRAGTTTA	1380
Db	4756	TAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAATGTTTAAGAAGTRAGTTTA	4815
Qy	1381	AGTTGTAGAGAGGCAGCCATGCATTTGGCATTTGAATACAATCTGGTGACTTGTCTGGCT	1440
Db	4816	AGTTGTAGAGAGGCAGCCATGCATTTGGCATTTGAATACAATCTGGTGACTTGTCTGGCT	4875
Qy	1441	GCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAG	1500
Db	4876	GCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAG	4935
Qy	1501	TCCTGCCTAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCA	1560
Db	4936	TCCTGCCTAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCA	4995
Qy	1561	AACCCAGAGC	1570
Db	4996	AACCCAGAGC	5005

RESULT 2

F351786S01

LOCUS	F351786S01	1000 bp	DNA	linear	ROD 23-AUG-2002
DEFINITION	Mus musculus sterolin-1 (Abcg5) gene, exon 1.				
ACCESSION	AF351786				
VERSION	AF351786.1 GI:18958385				
KEYWORDS	.				
SEGMENT	1 of 13				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1000)

AUTHORS Lu,K., Lee,M.-H., Yu,H., Zhou,Y., Sandell,S.A., Salen,G. and Patel,S.B.

TITLE Molecular cloning, genomic organization, genetic variations, and characterization of murine sterolin genes Abcg5 and Abcg8

JOURNAL J. Lipid Res. 43 (4), 565-578 (2002)

MEDLINE 21904563

PUBMED 11907139

REFERENCE 2 (bases 1 to 1000)

AUTHORS Lu,K., Zhou,Y., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

FEATURES

source Location/Qualifiers

1. .1000

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="129/Sv"

/db_xref="taxon:10090"

/chromosome="17"

/map="between Mit41 and Mit189"

/clone="329B11"

exon <359. .504

/gene="Abcg5"

/number=1

ORIGIN

Query Match 61.5%; Score 965.4; DB 10; Length 1000;

Best Local Similarity 99.0%; Pred. No. 2.3e-289;

Matches 991; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

Qy 64 GACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTGTCTTT 123

Db 1 GACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTGTCTTT 60

Qy 124 GCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTCTCCCAGCATTCCT 183

Db 61 TGTCCTTAGAGCTGGGGCACATGAGCCCTCCTCTGTGCCAGCCTTCTCCCAGCATTCCT 120

Qy 184 YTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACA 243

Db 121 CTCTGGCAAACAC-TCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACA 179

Qy 244 CTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTC 303

Db 180 CTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTC 239

Qy 304 ACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAA 363

Db 240 ACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAA 299

Qy 364 GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAGCCA 423

Db 300 GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAGCCA 359

Qy 424 TGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGT 483
 |||
 Db 360 TGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGT 419
 Qy 484 CTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTG 543
 |||
 Db 420 CTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTG 479
 Qy 544 TCCTGCATGTGTCTTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTC 603
 |||
 Db 480 TCCTGCATGTGTCTTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTC 539
 Qy 604 -TGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAG 662
 |||
 Db 540 TTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAG 599
 Qy 663 ATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCG 722
 |||
 Db 600 ATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCG 659
 Qy 723 CAGGGCAGAGCCTGGACATTAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTG 782
 |||
 Db 660 CAGGGCAGAGCCTGGACATTAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTG 719
 Qy 783 ATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCT 842
 |||
 Db 720 ATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCT 779
 Qy 843 CGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGT 902
 |||
 Db 780 CGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGT 839
 Qy 903 TCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATC 962
 |||
 Db 840 TCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATC 899
 Qy 963 AGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCCAA 1022
 |||
 Db 900 AGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCCAA 959
 Qy 1023 GGACAACAGAGTGGTACATAAGGCTAAACAGAGTTGTCAA 1063
 |||
 Db 960 GGACAACAGAGTGGTACATAAGGCTAAACAGAGTTGTCAA 1000

RESULT 3

AC120701/c

LOCUS AC120701 237445 bp DNA linear HTG 21-SEP-2002

DEFINITION Rattus norvegicus clone CH230-65H6, *** SEQUENCING IN PROGRESS ***,
 4 unordered pieces.

ACCESSION AC120701

VERSION AC120701.4 GI:23265381

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 237445)

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hagues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 237445)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 237445)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Sep 21, 2002 this sequence version replaced gi:21908396. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GXQV

Center clone name: CH230-65H6

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 209781 bases at least Q40

Consensus quality: 213033 bases at least Q30

Consensus quality: 214997 bases at least Q20

Estimated insert size: 233017; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 233866: contig of 233866 bp in length

* 233867 233966: gap of unknown length

* 233967 235011: contig of 1045 bp in length

* 235012 235111: gap of unknown length

* 235112 236137: contig of 1026 bp in length

* 236138 236237: gap of unknown length

* 236238 237445: contig of 1208 bp in length.

FEATURES Location/Qualifiers

source 1. .237445

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-65H6"

misc_feature 1. .1326

```

                                /note="wgs_end_extension
                                clone_end:T7"
misc_feature      8065. .8944
                                /note="clone_boundary
                                clone_end:T7
                                site:EcoRI
                                end_sequence:BH350813"
misc_feature      complement(232953. .233569)
                                /note="clone_boundary
                                clone_end:Sp6
                                site:EcoRI
                                end_sequence:BH350815"

```

ORIGIN

```

Query Match          55.3%;  Score 868.6;  DB 2;  Length 237445;
Best Local Similarity 81.6%;  Pred. No. 4.9e-259;
Matches 1181;  Conservative 3;  Mismatches 197;  Indels 66;  Gaps 13;

```

```

Qy          2 GAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGCC 61
              |||
Db    137687 GAAGCATCCTGGAGTACAGTCCCGTTCCACAGCTGGGTCTCCTCTTTGGTCTTCTCAGCC
137628

Qy          62 ATGACC-----AGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGG---- 109
              |||
Db    137627 ATGACCTGCGGTGTTGTGCCCTTTGTGTGGCTCCTGAGGCCTCCCCTGCTGTTGGCTAGG
137568

Qy          110 -----CTCTCTCTGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTG--- 160
              |||
Db    137567 CCAGGATTCTTTCTGTCTTTGCTCCTTAGAGCTAGGGCACTTGAGTCCTCCTTCCTGGCA
137508

Qy          161 CCAGCCTTTCTCCCAGCATTCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTTC 220
              |||
Db    137507 CCAGCCTTTCTCCCAGCATTCCTCTCTGGCAAGC-CCTCCTATAAACACACTGTGTGTTTC
137449

Qy          221 TGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG 280
              |||
Db    137448 TGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG
137389

Qy          281 CCAAATTGGTGAACGTGTTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCA 340
              |||
Db    137388 CCAAATCGGTGAACGTGTTGTCCTCACGAGGACTCTCGGGCTGGATAGGATCTGACAGGGCA
137329

Qy          341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
              |||
Db    137328 CTCCCATTTGGCTCCTCAGTTAAAGTTGCTCTGAAGCCAGACAGGACACCAGAGGATTCAC
137269

Qy          401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
              |||
Db    137268 TCACATTTGCTTCCCGCTGGCCATGAGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG
137209

```

Qy	461	AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCAC	520
Db	137208	AGGGCCTCACAAACAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTC	
	137149		
Qy	521	AGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCCTACAGCGTCAGGTAAGGGGACCT	580
Db	137148	AGAGGCTCGGCACAGCTTAGGTGTCTGAATGTGTCCTTCAGCGTCAGGTAAGGGGACC-	
	137090		
Qy	581	CCACAGCAAAAAGCTAGGCTCTC-----TGATTGCCTTTTCTGAATGGGTGG	627
Db	137089	CCACAGCGAAGAGCTAGGCTTCCCACCCTATCTGATGCCTTTTCACACCAAGGTGGGTGG	
	137030		
Qy	628	GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT	687
Db	137029	GTGGGCCTGTGGGCTTTGGGCTGCCTGTCTAGCAGATCAGGGTGGAAGTGGACAGTTCGT	
	136970		
Qy	688	AACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAC	747
Db	136969	TGCAACAGTGAGTGG-----CTCCTCCCCCTGCCAGAGCAGATCCTGAACATTGAAAC	
	136916		
Qy	748	ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCT-CCCCTTCCTTGAC	806
Db	136915	ACACCCTGCCTGAAGCCGC-TGCTGCTTCTCATAGATTTCTGCTCTACCCTTTCCTTGCC	
	136857		
Qy	807	TCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCA	866
Db	136856	TGGTCCATCACCTGCCCTCTGTAGATGGAGAAGGCTTGGGAAGTGGGGGTGCT-GGGACA	
	136798		
Qy	867	CAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGTGTG	926
Db	136797	CAAGGTGGAATGAACCCTGATGGAGGAATGCAGGGTTCACCTC-AGAATAAAGTGTACAT	
	136739		
Qy	927	CAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCT	986
Db	136738	GTTACCATCTCACAGCCAGACAGAGATCCAGTAATCAGAGCAGCCAAAGGGAGGCACGTT	
	136679		
Qy	987	CCTTGGGTGGTGGCCAACCTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGC	1046
Db	136678	CCTTGGGTGGTGGCCAACCTTGTCATTACACCTCCAAGGACCACATAGTGTGATGCAAGGC	
	136619		
Qy	1047	TAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAG-----	1101
Db	136618	TGAAATAGAGTTGTCTATCTGCACAGGAGACCTGGGATGGGGTTGGTCTGGGTGTGGGG	
	136559		
Qy	1102	-----CAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATT	1153

[illegible]

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hagues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 312858)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (24-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 312858)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Sep 23, 2002 this sequence version replaced gi:21738477. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRAX
Center clone name: CH230-359E1

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 241372 bases at least Q40
Consensus quality: 245333 bases at least Q30
Consensus quality: 248022 bases at least Q20
Estimated insert size: 276767; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 155105: contig of 155105 bp in length
* 155106 155205: gap of unknown length
* 155206 221765: contig of 66560 bp in length
* 221766 221865: gap of unknown length
* 221866 290378: contig of 68513 bp in length
* 290379 290478: gap of unknown length
* 290479 293724: contig of 3246 bp in length
* 293725 293824: gap of unknown length
* 293825 305790: contig of 11966 bp in length
* 305791 305890: gap of unknown length
* 305891 307341: contig of 1451 bp in length
* 307342 307441: gap of unknown length
* 307442 309768: contig of 2327 bp in length
* 309769 309868: gap of unknown length
* 309869 312858: contig of 2990 bp in length.

FEATURES	Location/Qualifiers
source	1. .312858 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-359E1"
misc_feature	159838. .161520 /note="wgs_contig"
misc_feature	166727. .168287 /note="wgs_contig"

```

misc_feature      190162. .191648
                  /note="wgs_contig"
misc_feature      234118. .235251
                  /note="wgs_contig"
misc_feature      290479. .292119
                  /note="wgs_contig"

```

ORIGIN

```

Query Match          55.3%;  Score 868.6;  DB 2;  Length 312858;
Best Local Similarity 81.6%;  Pred. No. 4.9e-259;
Matches 1181;  Conservative 3;  Mismatches 197;  Indels 66;  Gaps 13;

```

```

Qy      2  GAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTCTCAGCC 61
          |||||
Db      91501 GAAGCATCCTGGAGTACAGTCCCGTTCCACAGCTGGGTCTCCTCTTTGGTCTTCTCAGCC 91560

Qy      62  ATGACC-----AGTGTCTGTTTGTGCCCTTTGTGTGGCCTCCCGTGTGTTGGG---- 109
          |||||
Db      91561 ATGACCTGCGGTGTTGTGCCCTTTGTGTGGCTCCTGAGGCCTCCCGTGTGTTGGCTAGG 91620

Qy      110  -----CTCTCTCTGTCTTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTG--- 160
          |||
Db      91621 CCAGGATTCTTTCTGTCTTTTGCTCCTTAGAGCTAGGGCACTTGAGTCCTCCTTCCTGGCA 91680

Qy      161  CCAGCCTTTCTCCCAGCATTCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTC 220
          |||||
Db      91681 CCAGCCTTTCTCCCAGCATTCCTCTCTGGCAAGC-CCTCCTATAAACACACTGTGTGTTC 91739

Qy      221  TGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG 280
          |||||
Db      91740 TGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG 91799

Qy      281  CCAAATTGGTGAACGTGTTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCA 340
          |||||
Db      91800 CCAAATCGGTGAACGTGTTGTCTCACGAGGACTCTCGGGCTGGATAGGATCTGACAGGGCA 91859

Qy      341  CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
          |||||
Db      91860 CTCCCATTTGGCTCCTCAGTTAAAGTTGCTCTGAAGCCAGACAGGACACCAGAGGATTTCAC 91919

Qy      401  TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
          |
Db      91920 TCACATTTGCTTCCCGCTGGCCATGAGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 91979

Qy      461  AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGTCACGGGCAC 520
          |||||
Db      91980 AGGGCCTCACAACAACAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTC 92039

Qy      521  AGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTTACAGCGTCAGGTAAGGGGACCT 580
          |||||
Db      92040 AGAGGCTCGGCACAGCTTAGGTGTCTGAATGTGTCTTACAGCGTCAGGTAAGGGGACC- 92098

Qy      581  CCACAGCAAAAAGCTAGGCTCTC-----TGATTGCCTTTTCTGAATGGGTGG 627
          |||||
Db      92099 CCACAGCGAAGAGCTAGGCTTCCACCCCTATCTGATGCCTTTTCACACCAAGGTGGGTGG 92158

Qy      628  GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687

```

Db	92159	 GTGGGCCTGTGGGCTTTGGGCTGCCTGTCTAGCAGATCAGGGTGGAAAGTGGACAGTTTCGT	92218
Qy	688	AACAACAGTGAGTCGTTCCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC	747
Db	92219	 TGCAACAGTGAGTGG-----CTCCTCCCCCTGCCAGAGCAGATCCTGAACATTGAAAC	92272
Qy	748	ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCT-CCCCTTCCTTGAC	806
Db	92273	 ACACCCTGCCTGAAGCCGC-TGCTGCTTCTCATAGATTTCTGCTCTACCCTTTCCTTGGC	92331
Qy	807	TCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCA	866
Db	92332	 TGGTCCATCACCTGCCCTCTGTAGATGGAGAAGGCTTGGGAAGTGGGGGTGCT-GGGACA	92390
Qy	867	CAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGTGTG	926
Db	92391	 CAAGGTGGAATGAACCCTGATGGAGGAATGCAGGGTTCACCTC-AGAATAAAGTGTACAT	92449
Qy	927	CAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCT	986
Db	92450	 GTTACCATCTCACAGCCAGACAGAGATCCAGTAATCAGAGCAGCCAAAGGGAGGCACGTT	92509
Qy	987	CCTTGGGTGGTGGCCAACTTGTCAATTATACCTCCAAGGACAACAGAGTGGTACATAAGGC	1046
Db	92510	 CCTTGGGTGGTGGCCAACTTGTCAATTACACCTCCAAGGACCACATAGTGTGATGCAAGGC	92569
Qy	1047	TAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAG-----	1101
Db	92570	 TGAAATAGAGTTGTCACTTGCACAGGAGGACCTGGGATGGGGTTGGTCTGGGTGTGGGG	92629
Qy	1102	-----CAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATT	1153
Db	92630	 CTGGGAAACAGGGGTCTGGCACCTTCAAGGGTCTACTCTGCCTTTTGTTCATGTGGGAT	92689
Qy	1154	TCCTTTTAAAGCAACCGTGTGCGGCCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGG	1213
Db	92690	 TTCTTTTAAAGCAACCGTGTGCGGCCCTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGG	92749
Qy	1214	ACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCT	1273
Db	92750	 ACAGGAAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGACCATGTGCATCT	92809
Qy	1274	TAGGCAGCTCAGGTAAGTGCCTGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCT	1333
Db	92810	 TAGGTAGCTCAGGTAAGCGCCT-----CGAGGGGTCTGCACTTGTAAAGCAGACTCT	92862
Qy	1334	GGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTAAGTTGTAGAGAGG	1393
Db	92863	 GGGAGGCTGGGGCTCGGTCTAAGCTCGGTGTTTAAGAAATGAGTTTAATTGGGAGGGGAA	92922
Qy	1394	CAGCCAT 1400	
Db	92923	 CACCCAT 92929	

AY145899
 LOCUS AY145899 40929 bp DNA linear ROD 12-NOV-2002
 DEFINITION Rattus norvegicus sterolin 2 (Abcg8) and sterolin 1 (Abcg5) genes, complete cds.
 ACCESSION AY145899
 VERSION AY145899.1 GI:24935208
 KEYWORDS .
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 40929)
 AUTHORS Yu, H., Lu, K., Lee, M., Pandit, B. and Patel, S.B.
 TITLE The rat Abcg5 and Abcg8: characterization, chromosomal assignment and genetic variation in sitosterolemic rats
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 40929)
 AUTHORS Yu, H., Lu, K., Lee, M., Pandit, B. and Patel, S.B.
 TITLE Direct Submission
 JOURNAL Submitted (29-AUG-2002) Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty Street, STR 541, Charleston, SC 29403, USA
 FEATURES
 source Location/Qualifiers
 1. .40929
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 gene complement(<4136. .>20831)
 /gene="Abcg8"
 mRNA complement(join(<4136. .4273,4361. .4488,5693. .5960, 6513. .6589,6754. .6953,8189. .8269,8350. .8512,10772. .11041, 11129. .11261,11647. .11885,15513. .15669,17473. .17574, 20769. .>20831))
 /gene="Abcg8"
 /product="sterolin 2"
 CDS complement(join(4136. .4273,4361. .4488,5693. .5960, 6513. .6589,6754. .6953,8189. .8269,8350. .8512,10772. .11041, 11129. .11261,11647. .11885,15513. .15669,17473. .17574, 20769. .20831))
 /gene="Abcg8"
 /note="ATP-binding cassette sub-family G (WHITE) member 8"
 /codon_start=1
 /product="sterolin 2"
 /protein_id="AAN64276.1"
 /db_xref="GI:24935210"
 /translation="MAEKTKEETQLWNGTVLQDASSLQDSVFSSSEDNSLYFTYSGQS
 NTLEVRDLTYQVDMASQVPWFQQLAQFKLPWRSRGSQDSWDLGIRNLSFKVRSQMLA
 IIGSAGCGRATLLDVITGRDHGGMKSGQIWIWGPSTPQLIQKCAHVRRQQDQLLPN
 LTVRETTLTFIAQMRLPKTFSSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
 RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVRTLSRLAKGNRLVLISLHQPRSDI
 FRLFDLVLLMTSGTPIIYLGVAQHMQYFTSIGYPCPRYSNPADFYVDLTSIDRRSKEQ
 EVATMEKARLLAALFLEKVQGFDDFLWKAELSLDTGTAYVSQTLTQDTNCGTAAELP
 GMIQQFTTLIRRQISNDFRDLPTLFIHGAEACLSLIIGFLYYGHADKPLSFMDMAAL

Qy	221	TGCCTATTGTCGAGATAAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG	280
Db	21009	TGCCTATTGTCGAGATAAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG	21068
Qy	281	CCAAATTGGTGAACTGTTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCA	340
Db	21069	CCAAATCGGTGAACTGTTGTCTCACGAGGACTCTCGGGCTGGATAGGATCTGACAGGGCA	21128
Qy	341	CTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTAC	400
Db	21129	CTCCCATTGGCTCCTCAGTTAAAGTTGCTCTGAAGCCAGACAGGACACCAGAGGATTAC	21188
Qy	401	TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG	460
Db	21189	TCACATTTGCTTCCCGCTGGCCATGAGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG	21248
Qy	461	AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGTACAGGGCAC	520
Db	21249	AGGGCCTCACAACAACAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTC	21308
Qy	521	AGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCCTACAGCGTCAGGTAAGGGGACCT	580
Db	21309	AGAGGCTCGGCACAGCTTAGGTGTCTGAATGTGTCCTTCAGCGTCAGGTAAGGG--ACC	21366
Qy	581	CCACAGCAAAAAGCTAGGCTCTC-----TGATTGCCTTTTCTGAATGGGTGG	627
Db	21367	CCACAGCGAAGAGCTAGGCTTCCCACCCTATCTGATGCCTTTTCACACCAAGGTGGGTGG	21426
Qy	628	GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT	687
Db	21427	GTGGGCCTGTGGGCTTTGGGCTGCCTGTCTAGCAGATCAGGGTGGAAGTGGACAGTTCGT	21486
Qy	688	AACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC	747
Db	21487	TGCAACAGTGAGTGG-----CTCCTCCCCCTGCCCAGAGCAGATCCTGAACATTGAAAC	21540
Qy	748	ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCT-CCCCTTCCTTGAC	806
Db	21541	ACACCCTGCCTGAAGCCGC-TGCTGCTTCTCATAGATTTCTGCTCTACCCTTTCCTTGGC	21599
Qy	807	TCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCA	866
Db	21600	TGGTCCATCACCTGCCCTCTGTAGATGGAGAAGGCTTGGGAAGTGGGGGTGCT-GGGACA	21658
Qy	867	CAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGTGTG	926
Db	21659	CAAGGTGGAATGAACCCTGATGGAGGAATGCAGGGTTCACCTC-AGAATAAAGTATACAT	21717
Qy	927	CAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCT	986
Db	21718	GTTACCATCTCACAGCCAGACAGAGATCCAGTAATCAGAGCAGCCAAAGGGAGGCACGTT	21777
Qy	987	CCTTGGGTGGTGGCCAACCTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGC	1046
Db	21778	CCTTGGGTGGTGGCCAACCTTGTCATTACACCTCCAAGGACCACATAGTGTGATGCAAGGC	21837

Qy	1047	TAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATAGGGGTAGGGCTGGGAG-----	1101
Db	21838	TGAAATAGAGTTGTTCATCTTGACAGGAGGACCTGGAATGGGGTTGGTCTGGGTGTGGGG	21897
Qy	1102	-----CAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATT	1153
Db	21898	CTGGGAAACAGGGTCTGGCACCTTCAAGGGTCTACTCTGCCTTTTGTTCATGTGGATT	21957
Qy	1154	TCCTTTAAAGCAACCGTGTCTGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGG	1213
Db	21958	TCCTTTAAAGCAACCGTGTCTGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGG	22017
Qy	1214	ACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCT	1273
Db	22018	ACAGGAAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGACCATGTGCATCT	22077
Qy	1274	TAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCT	1333
Db	22078	TAGGTAGCTCAGGTAAGCGCCT-----CGAGGGGTCTCTGCACTTGTAAAGGCAGACTCT	22130
Qy	1334	GGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTAAGTTGTAGAGAGG	1393
Db	22131	GGGAGGCTGGGGCTCGGTCTAAGCTCGGTGTTTAAGAAATGAGTTTAATTGGGAGGGGAA	22190
Qy	1394	CAGCCAT 1400	
Db	22191	CACCCAT 22197	


```

gene      /db_xref="taxon:10090"
          complement(<1. .>146)
          /gene="Abcg5"
mRNA      complement(<1. .>146)
          /gene="Abcg5"
          /product="sterolin 1"
CDS       complement(<1. .146)
          /gene="Abcg5"
          /codon_start=1
          /product="sterolin 1"
          /protein_id="AAK85390.1"
          /db_xref="GI:15150322"
          /translation="MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHV
          SYSV"
exon      complement(<1. .>146)
          /gene="Abcg5"
          /number=1
misc_feature 147. .504
          /note="contains 5'UTR and promoter regions for ABCG5 and
          ABCG8"
gene      <505. .>567
          /gene="Abcg8"
mRNA      <505. .>567
          /gene="Abcg8"
          /product="sterolin 2"
CDS       505. .>567
          /gene="Abcg8"
          /codon_start=1
          /product="sterolin 2"
          /protein_id="AAK85391.1"
          /db_xref="GI:15150323"
          /translation="MAEKTKEETQLWNGTVLQDAS"
exon      <505. .>567
          /gene="Abcg8"
          /number=1

```

ORIGIN

```

Query Match      35.4%;  Score 555.6;  DB 10;  Length 567;
Best Local Similarity 99.6%;  Pred. No. 2.1e-161;
Matches 566;  Conservative 1;  Mismatches 0;  Indels 1;  Gaps 1;

```

```

Qy      1  CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC  60
         |||
Db      567 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC  508
         |||
Qy      61  CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC  120
         |||
Db      507  CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC  448
         |||
Qy      121 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT  180
         |||
Db      447 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT  388
         |||
Qy      181 CCTYTCTGGCAAACACTTCCATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG  240
         |||:
Db      387 CCTCTCTGGCAAACAC-TCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG  329

```

```

Qy      241 ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      328 ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTAT 269

Qy      301 CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      268 CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT 209

Qy      361 AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC TTGCATTTGCTTCCTGCTAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      208 AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC TTGCATTTGCTTCCTGCTAG 149

Qy      421 CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      148 CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG 89

Qy      481 GGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTACAGGGGCACAGAGGCTCGGCACAGCTTAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88  GGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTACAGGGGCACAGAGGCTCGGCACAGCTTAG 29

Qy      541 GTGTCCTGCATGTGTCCTACAGCGTCAG 568
          ||||||||||||||||||||||||||||
Db      28  GTGTCCTGCATGTGTCCTACAGCGTCAG 1

```

RESULT 7

AF404109/c

LOCUS AF404109 588 bp DNA linear ROD 14-AUG-2001

DEFINITION Rattus norvegicus sterolin 1 (Abcg5) and sterolin 2 (Abcg8) genes, partial cds.

ACCESSION AF404109

VERSION AF404109.1 GI:15150324

KEYWORDS .

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 588)

AUTHORS Lu,K., Lee,M. and Patel,S.B.

TITLE Molecular cloning, genomic structure and characterization of novel murine ABC genes Abcg5 and Abcg8

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 588)

AUTHORS Lu,K., Lee,M. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (11-JUL-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403

FEATURES Location/Qualifiers

```

source          1..588
                  /organism="Rattus norvegicus"
                  /mol_type="genomic DNA"
                  /strain="Sprague-Dawley"
                  /db_xref="taxon:10116"
gene            complement(<1..>146)
                  /gene="Abcg5"

```

```

mRNA      complement(<1. .>146)
          /gene="Abcg5"
          /product="sterolin 1"
CDS       complement(<1. .146)
          /gene="Abcg5"
          /codon_start=1
          /product="sterolin 1"
          /protein_id="AAK85392.1"
          /db_xref="GI:15150325"
          /translation="MGELPFLSPEGARGPHNNRGSQSSLEEGSVTGSEARHSLGVLNV
          SFSV"
exon      <1. .>146
          /number=1
misc_feature 147. .525
          /note="contains 5'UTR and promoter regions for ABCG5 and
          ABCG8"
gene      <526. .>588
          /gene="Abcg8"
mRNA      <526. .>588
          /gene="Abcg8"
          /product="sterolin 2"
CDS       526. .>588
          /gene="Abcg8"
          /codon_start=1
          /product="sterolin 2"
          /protein_id="AAK85393.1"
          /db_xref="GI:15150326"
          /translation="MAQT'TKEETQLWNGTVLQDAS"
exon      <526. .>588
          /gene="Abcg8"
          /number=1

```

ORIGIN

```

Query Match      26.3%;  Score 412.4;  DB 10;  Length 588;
Best Local Similarity 86.4%;  Pred. No. 1.1e-116;
Matches 508;  Conservative 1;  Mismatches 57;  Indels 22;  Gaps 4;

```

```

Qy      2 GAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGCC 61
        ||||||| ||||||| ||||||| ||||||| ||| |||
Db      587 GAAGCATCCTGGAGTACAGTCCCGTTCACAGCTGGGTCTCCTCTTTGGTCGTCTGAGCC 528

Qy      62 ATGACC-----AGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTT----- 106
        ||||| ||| ||||| | || |||||||||||||
Db      527 ATGACCTGCGGTGTTGTGCCCTTTGTGTGGCTCCTGAGGCCTCCCCTGCTGTTGGCTAGG 468

Qy      107 ---GGGCTCTCTCTGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTG--- 160
        || || ||||||||||||||||||| ||||| ||| ||||| ||
Db      467 CCAGGATTCTTTCTGTCTTTGCTCCTTAGAGCTAGGGCACTTGAATCCTTCCTGGCA 408

Qy      161 CCAGCCTTTCTCCCAGCATTCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTC 220
        |||||||||||||||||||:|||||| | | ||||||||||| |||||||
Db      407 CCAGCCTTTCTCCCAGCATTCCTCTCTGGCAAGC-CCTCCTATAAACACACTGTGTGTTC 349

Qy      221 TGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG 280
        ||||||||||||||||||| |||||||||||||||||||
Db      348 TGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGG 289

```

Qy 281 CCAAATTGGTGAAC TGTATCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCA 340
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 288 CCAAATCGGTGAAC TGTGTCTCACGAGGACTCTCGGGCTGGATAGGATCTGACAGGGCA 229

Qy 341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 228 CTCCCATTTGGCTCCTCAGTTAAAGTTGCTCTGAAGCCAGACAGGACACCAGAGGATTCAC 169

Qy 401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 168 TCACATTTGCTTCCCGCTGGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 109

Qy 461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGTCACGGGCAC 520
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 108 AGGGCCTCACAACAACAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTC 49

Qy 521 AGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAG 568
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 48 AGAGGCTCGGCACAGCTTAGGTGTCCTGAATGTGTCCTTCAGCGTCAG 1

RESULT 8

F351786S02

LOCUS F351786S02 463 bp DNA linear ROD 23-AUG-2002

DEFINITION Mus musculus sterolin-1 (Abcg5) gene, exon 2.

ACCESSION AF351787

VERSION AF351787.1 GI:18958386

KEYWORDS .

SEGMENT 2 of 13

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 463)

AUTHORS Lu,K., Lee,M.-H., Yu,H., Zhou,Y., Sandell,S.A., Salen,G. and
 Patel,S.B.

TITLE Molecular cloning, genomic organization, genetic variations, and
 characterization of murine sterolin genes Abcg5 and Abcg8

JOURNAL J. Lipid Res. 43 (4), 565-578 (2002)

MEDLINE 21904563

PUBMED 11907139

REFERENCE 2 (bases 1 to 463)

AUTHORS Lu,K., Zhou,Y., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers

source 1. .463
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="129/Sv"
 /db_xref="taxon:10090"
 /chromosome="17"
 /map="between Mit41 and Mit189"
 /clone="329B11"
 exon 101. .222

ORIGIN

Qy	1064	CCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGG	1123
Db	1	CCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGG	60
Qy	1124	ACCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTTAAAGCAACCGTGTCTGGGCCTTGGT	1183
Db	61	ACCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTTAAAGCAACCGTGTCTGGGCCTTGGT	120
Qy	1184	GGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGT	1243
Db	121	GGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGT	180
Qy	1244	ACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSC	1303
Db	181	ACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGGC	240
Qy	1304	SGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAATG	1363
Db	241	CGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCTGTCTAAGCACAATG	300
Qy	1364	TTTAAGAACTTAGTTTAAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATC	1423
Db	301	CTTAAGAACTTAGTTTAAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATC	360
Qy	1424	TGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTT	1472
Db	361	TGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATATT	409

F351799S01/c

LOCUS	F351799S01	1314 bp	DNA	linear	ROD 23-AUG-2002
DEFINITION	Mus musculus sterolin 2 (Abcg8) gene, exon 1.				
ACCESSION	AF351799				
VERSION	AF351799.1 GI:18996437				
KEYWORDS	.				
SEGMENT	1 of 13				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 1314)				
AUTHORS	Lu,K., Lee,M.-H., Yu,H., Zhou,Y., Sandell,S.A., Salen,G. and Patel,S.B.				
TITLE	Molecular cloning, genomic organization, genetic variations, and characterization of murine sterolin genes Abcg5 and Abcg8				
JOURNAL	J. Lipid Res. 43 (4), 565-578 (2002)				
MEDLINE	21904563				
PUBMED	11907139				

REFERENCE 2 (bases 1 to 1314)
 AUTHORS Lu,K., Zhou,Y., Lee,M.-H. and Patel,S.B.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St., STB 541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers
 source 1. .1314
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="129/Sv"
 /db_xref="taxon:10090"
 /chromosome="17"
 /map="between Mit41 and Mit189"
 /clone="329B11"
 exon <359. .421
 /gene="Abcg8"
 /number=1

ORIGIN

Query Match 25.4%; Score 398.4; DB 10; Length 1314;
 Best Local Similarity 97.9%; Pred. No. 2.6e-112;
 Matches 413; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 362
 Qy 61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 361 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 302
 Qy 121 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT 180
 ||| | | ||||||||||| ||||||||||||||||||||||||||||||||||||
 Db 301 TTTTGCTCCTTAGAGCTGGGGCACATGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT 242
 Qy 181 CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCTGAGATAAGG 240
 |||:||||||||| ||||||||||||||||||||||||||||||||||||
 Db 241 CCTCTCTGGCAAACAC-TCCTATAAACACACCGTGTGTTCTGCCTATTGTCTGAGATAAGG 183
 Qy 241 ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTAT 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 182 ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTAT 123
 Qy 301 CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 122 CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT 63
 Qy 361 AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAG 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 62 AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAG 3
 Qy 421 CC 422
 ||
 Db 2 CC 1

RESULT 10
 AX685738
 LOCUS AX685738 359 bp DNA linear PAT 29-MAR-2003
 DEFINITION Sequence 10 from Patent WO02081691.
 ACCESSION AX685738
 VERSION AX685738.1 GI:29371747
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.
 TITLE Abcg5 and abcg8: compositions and methods of use
 JOURNAL Patent: WO 02081691-A 10 17-OCT-2002;
 Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
 (US)
 FEATURES Location/Qualifiers
 source 1. .359
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="sequence between ABCG5 and ABCG8 containing control
 sequences (bidirectional promoter)"
 ORIGIN
 Query Match 22.8%; Score 358.6; DB 6; Length 359;
 Best Local Similarity 100.0%; Pred. No. 6.6e-100;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 64 GACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTGTCTTT 123
 |||||
 Db 1 GACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTGTCTTT 60
 Qy 124 GCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCT 183
 |||||
 Db 61 GCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCT 120
 Qy 184 YTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACA 243
 |||||
 Db 121 YTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACA 180
 Qy 244 CTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTC 303
 |||||
 Db 181 CTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTC 240
 Qy 304 ACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAA 363
 |||||
 Db 241 ACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAA 300
 Qy 364 GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTGCATTTGCTTCCTGCTAGCC 422
 |||||
 Db 301 GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTGCATTTGCTTCCTGCTAGCC 359

RESULT 11
 AC146466/c

LOCUS AC146466 185045 bp DNA linear HTG 15-AUG-2003
 DEFINITION Callithrix jacchus clone CH259-274K20, WORKING DRAFT SEQUENCE, 3 ordered pieces.
 ACCESSION AC146466
 VERSION AC146466.1 GI:33667132
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
 SOURCE Callithrix jacchus (white-tufted-ear marmoset)
 ORGANISM Callithrix jacchus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
 REFERENCE 1 (bases 1 to 185045)
 AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R., Peng, Z., Malinov, I. and Rubin, E.M.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 185045)
 AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R., Peng, Z., Malinov, I. and Rubin, E.M.
 TITLE Direct Submission
 JOURNAL Submitted (15-AUG-2003) Genome Sciences, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
 COMMENT
 Sequence Produced by Berkeley PGA
 Web site: <http://pga.lbl.gov>
 Center Code: PGABERK
 Center Project Name: J027
 Bac Clone Name: CH259-274K20

 This sequence has been compared to sequences of other species using Vista (<http://www-gsd.lbl.gov/VISTA>). The results can be viewed at:
http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=ABCG5

 The order-orientation of the draft sequence was accomplished by using:
 Avid (<http://baboon.math.berkeley.edu/mavid>),
 Lagan (<http://lagan.stanford.edu/>) and paired end information.

 Funding agent: Programs for Genomic Applications (NHLBI)

 Summary Statistics:
 Sequencing vector: Plasmid; pUC18
 Chemistry: Dye-terminator Big Dye
 Assembly program: Phrap version 0.990329.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 49109: contig of 49109 bp in length
 * 49110 49209: gap of unknown length
 * 49210 57420: contig of 8211 bp in length


```

*      57421      57520: gap of unknown length
*      57521      185045: contig of 127525 bp in length.
FEATURES
    source          Location/Qualifiers
                    1. .185045
                    /organism="Callithrix jacchus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9483"
                    /clone="CH259-274K20"

```

ORIGIN

```

Query Match          19.1%;  Score 299.4;  DB 2;  Length 185045;
Best Local Similarity 57.7%;  Pred. No. 2.6e-81;
Matches 824;  Conservative 2;  Mismatches 513;  Indels 88;  Gaps 13;

```

```

Qy      120  CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
          || || | | | |||| | | |||| | | | | | | | | | | | | | | | | |
Db      116200 CTCTGTTTCCTGGAGCAGGGACACCTCAGCCTCCTGCCCTGGGCCCGGCTCTCCCAGCAT
116141

Qy      180  TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTTCGAGATAAG 239
          ||||:|||||| | | ||| | |||| | |||| | |||| | | | | | | | |
Db      116140 TCCTCTCTGGCAAGCCCA-CCTACAAACACAT-GTGTGTTCTGCCCTCTCTCAAGATAAG
116083

Qy      240  GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTTA 299
          ||| | ||||| | | | | | | | | | | | | | | | | | | | | | |
Db      116082 GACGCGCTGGCTAAAGGTACATCAGATAACGGCCTTCTTGGCCAAAGTCCCAGTCCCTGCCA
116023

Qy      300  TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT 359
          || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      116022 TCCTGAGGGACTCTGGGGTCAGGTGGAGCTGGCAGGGCAGTCTGCCACTGGCTCCCCAAC
115963

Qy      360  TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCATTTGCTTCCTGCTA 419
          | ||| | ||| | | |||| | | |||| | | | | | | | | | | | |
Db      115962 TGCAGCCACTCTGAGGAGGGTCAGGCTACCAGAAAATCTGCCCAGCTTTGCTGCCCGTTG
115903

Qy      420  GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA 479
          ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db      115902 GCCATGGGTGACCTTCCATCTTTGACCCCGGAGGGTCCATAGGACTCCAGGTAAACAGA
115843

Qy      480  GGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTA 539
          || || | ||||| | | | | | | | | | | | | | | | | | | |
Db      115842 GGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACTGCACCTGAGCCT---CACAGTCTG
115786

Qy      540  GGTGTCCTGCATGTGTCTTACAGCGTCAGGTAAGG-----GGACCTCCA 583
          || |||| | ||| | |||| | | |||| | | | | | | | | | | | |
Db      115785 GGCATCCTCCATGCCTCCTACAGCATCAGGTAAGGCAGAGACCTTGCTGCTGCTCCTCCC
115726

Qy      584  CAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTT 643
          ||| | | | | ||| | | | | | | | | | | | | | | | | |

```

Db 115725 CAGGAGCACGGGGCCCTCTGCTGCCTTTTTTCACTCTTGAGCTGCCTGGCTGGAGACTTT
 115666

Qy 644 TGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGT 703
 ||| | || | ||| ||||| || || || || ||

Db 115665 GGGGCTCCCTCTTCAGTGGATCAGGTGGAGAGAAGAGAGGGGGAAGGG-----
 115618

Qy 704 TCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGC 763
 | | | | || || | | || | | ||||| |

Db 115617 -----CTGCACTGGGAAATAGGGAGCAACAGTAAATGGCCCCTCCCCCTGCCAGGGA
 115565

Qy 764 CGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTTGACTCGC--CCACCACCTGT 821
 | | | | || | || | || | || | || | || | || | ||

Db 115564 AGGGCCTAGGTATAAACAAGTTGAGCTGTGCCCTGCCTACCCTAGTGTCTACCCTTGC
 115505

Qy 822 CCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAAC 881
 ||| || ||||| || | |||| | | | | |

Db 115504 CCTCTGCAGATGGAGAGAATCTGGGGCTGGGGAGCTGGGAATAAAGGAGTCTTGAATCC
 115445

Qy 882 A-CTGCTGAAGGAATGCAGGGTTCCTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCT 940
 | || ||| | | | | || | || | | | | | | |

Db 115444 AGGTGACGAATGTAGGGACAACCACCTCCAGACAAATGGGCAGGACATTTGGAGCAGCT
 115385

Qy 941 CCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGC 1000
 || ||| ||| | || ||| ||| ||| ||

Db 115384 CCAGCACAGGCCCCC-----TCCCTAGGTGACAGACAGCCTCAGTCGCTACCTGC
 115335

Qy 1001 CAACTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGT 1060
 || || | | | | | || | || | | | | | |

Db 115334 CAGGTTCTACAGA-----GAAGGATGCCGAGGCTGAAACACGTTAGGAGCCTGTCTGA
 115282

Qy 1061 CAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCC 1120
 | | | || | || || | |||| | || | |||||

Db 115281 AGATAACTGGGGTGGGACACAGGTGGGATCAATGCTGGGGACCCAGGTGTAGCCCCCTTC
 115222

Qy 1121 AGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCCCTTTAAAGCAACCGTGTGCGGCCCTT 1180
 ||| ||| | ||||| || | ||||| ||||| |||||

Db 115221 AGGGCCCCATGCTGCCTTTGCTTTCCTGGGATTTCCCTTTAAAGCCACCGTGTGGGGCCCT
 115162

Qy 1181 GGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCT 1240
 |||| ||||| || ||||| |||| | ||||| ||||| |||||

Db 115161 GGTGGGACATCACATCTTGCCAGCGACAGTGGACCAGGCAGATCCTCAAAGACGTCTCCT
 115102

Qy 1241 TGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGG 1300
 |||| | ||||| |||| | ||||| |||| | |||||

Db 115101 TGTACGTGGAGAGTGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTAAGCTTGGGAAGA
 115042

```

Qy          1301  GSCSGGGGCTCCTGTACTTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCAC 1360
              :                               |   ||||| |||||   ||| | |
Db          115041 AG-----GATTTTAAAAAGGCTTTGGCTTGAGTTAAACTCC
115006

Qy          1361  ATGTTTAAAGAAGTRAGTTTAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTTGAATACA 1420
              |  || |||||   || |  ||||| ||  | |||   |||  ||| |||| |
Db          115005 ACCTTAAAGAA-ACAGATACAGTTGTAGCAAGAAAACACAGGTTTGATATTAGAATGAA
114947

Qy          1421  ATCTGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAA 1480
              ||||  |||  ||||| |||  ||||| ||||  ||||| ||||| || || |
Db          114946 ATCTAATGA--TGTCTGACTGTGAATAGAACCTGCTACCAATGTGAAATCTATAGAAAGA
114889

Qy          1481  TCCCTGGAAAGAGTGGAAAGTCCTGCCTAACACGTAAGTGCCTTCTT 1527
              |  ||||| |||||   ||| ||||| ||||| |||  ||  ||| |
Db          114888 T-CCTGGAAAGAGTATAAAATCCTGCCTAACATGTACATGAATTCAT 114843

```

using:
Avid (<http://baboon.math.berkeley.edu/mavid>),
Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

Summary Statistics:

Sequencing vector: Plasmid; pUC18

Chemistry: Dye-terminator Big Dye

Assembly program: Phrap version 0.990329.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 32150: contig of 32150 bp in length
* 32151 32250: gap of unknown length
* 32251 56222: contig of 23972 bp in length
* 56223 56322: gap of unknown length
* 56323 173105: contig of 116783 bp in length
* 173106 173205: gap of unknown length
* 173206 178016: contig of 4811 bp in length.

FEATURES
source Location/Qualifiers
1. .178016
/organism="Aotus nancymae"
/mol_type="genomic DNA"
/db_xref="taxon:37293"
/clone="CH258-323A5"

ORIGIN

Query Match 19.0%; Score 298.8; DB 2; Length 178016;
Best Local Similarity 56.2%; Pred. No. 4.1e-81;
Matches 869; Conservative 2; Mismatches 584; Indels 91; Gaps 13;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTCTCAGC 60
| | ||||| | | | | | | | | | | | | | | | | |
Db 90827 CAAGGCATCCTGGGGAGTGGCCCTTTTCGGCAGCCCTCTCTCCTCGGTGGCCTCCCAGC 90768

Qy 61 CAT-----GACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGC 110
| | | | | | | | | | | | | | | | | | | | | |
Db 90767 CATGGGGCCACAGGTCTGTGCCGTTGGGCTCAGCTCTTAGACCGGGGCTGCTGCCTGTC 90708

Qy 111 TCTCTCTGT-----CTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCA 163
| | | | | | | | | | | | | | | | | | | | | |
Db 90707 AGGGCCAGTGTCTTCGCTCTGTTTCTGGAGCAGGGACACATCAGCCTCCTGTCTGGGC 90648

Qy 164 GCCTTTCTCCAGCATTCCTYTCTGGCAAACACTTCCTATAAAACACACCGTGTGTTCTGC 223
| | | | | | | | | | | | | | | | | | | | | |
Db 90647 CCGGCTCTCCAGCATTCCTCTCTGGCAAG----TCCACCTAAAAACACGTGTGTTCTGC 90592

Qy 224 CTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCA 283
| | | | | | | | | | | | | | | | | | | | | |
Db 90591 CCTCTCTCAAGATAAGGACGCGCTGGCTAAAGGTACATCAGATAACGGCCTCCTTGCCA 90532

Qy	284	AATTGGTGAACGTGTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTC	343
Db	90531	AGTTCAGTCCTGCCATCCTGAGGGACTCCGGGGTCAGGTGGAGCTGGCAGGGCAGTCTG	90472
Qy	344	CCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTG	403
Db	90471	CCACTGGCTTCCCAACTGCAGCCACTCCGAGGAGGGTCAGGCTACCAGAAAATCTGCCCA	90412
Qy	404	CATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGG	463
Db	90411	GCTTTGCTGCCCGTTGGCCATGGGTGACCTTCCATCTTTAACCCCTCGGAGGGTCCATAGG	90352
Qy	464	GCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGA	523
Db	90351	ACTCCAGGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACTGCACCTGA	90292
Qy	524	GGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCA	583
Db	90291	GCCT---CACAGTCTGGGCATCCTCCATGCCTCTACAGCGTCAGGTAAGGCAGAGCCCT	90235
Qy	584	CAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTT	643
Db	90234	TGCTGCTGCTC---CTCCCCAGGAGCACGTTCACTCTGAGCTGCCTGGCTGGGGACTT	90178
Qy	644	TGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGT	703
Db	90177	TGGGCTCCCTCTTCAGTGGATCGGGTGGAGAGAAGAGAGCGGGGA-----	90133
Qy	704	TCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGC	763
Db	90132	----GGGCTGCACTGGGAAATGGGAGCAACAGTAAATGGCCCCCTCCCCCTGCCAGGGA	90077
Qy	764	CGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTTGAC--TCGCCCACCACCTGT	821
Db	90076	AGGGTCTGGTATAAAACAAAGTTGCAGCTGTGCCCTGCCTACCCAGTGTCTACCACTTGC	90017
Qy	822	CCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAAC	881
Db	90016	CCTCTGCAGATGGAGAGAATCTGGGGAATCGGGG-GCTGGGAATGCAAGAGTCTTGAAT	89958
Qy	882	ACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTC	941
Db	89957	CCAGGTGACGAA-----TGCAGGGACAACCACTTCCCAGACAAAATGGGCAG	89912
Qy	942	CCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCC	1001
Db	89911	GACATTCTGGAGCAGCTCCAGCACAGGCCCCCTCCCTAGGTGACAGACAGCCTCGGTAGCT	89852
Qy	1002	AACCTGTTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTC	1061
Db	89851	ACCTGCCAGGTTCTACAGAGGAGGATGCCGAGGCTGAAACACGTTAGGAGCCTGTCTGAA	89792
Qy	1062	AACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCA	1121
Db	89791	GATAACTGGGGTGGGACACAGGTGGGATCAACGCTGGGGACCTGGGTGTAGCCCCCTCCA	89732

Qy 1122 GGACCTACTCTGCCTTTGCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTGGGCCCTTG 1181
 || ||| | ||||| | ||||| ||||| ||||| ||||| ||
 Db 89731 GGGCCCCATGCTGCCTTTGCCCTTCCTGGGATTTTCCTTTAAAGCCACCGTGTGGGGCCCTG 89672

Qy 1182 GTGGAACATCAAATCATGCCAGCAGAAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTT 1241
 |||| ||||| ||| ||| || ||||| ||||| ||||| ||||| ||||| ||
 Db 89671 GTGGGACATCACATCTTGTGCTGGCGACAGTGGACCAGGAGATCCTCAAAGACGTCTCCTT 89612

Qy 1242 GTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGG 1301
 |||| | ||||| || ||||| ||||| ||||| ||||| ||||| || ||
 Db 89611 GTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTAAGCTTGGGAAGAA 89552

Qy 1302 SCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAA 1361
 : | ||||| ||||| ||| ||| ||
 Db 89551 G-----GATTTTAAAAAGGCTTTGGCTTGAGTTAAACTCCA 89516

Qy 1362 TGTTTAAAGAAGTRAGTTTAAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTTGAATACAA 1421
 | |||| | | | ||||| || |||| | ||| ||| ||| ||
 Db 89515 CCCTGAAGAA-ACAGATACAGTTGTAGCAAGAAAGCCACAGGTTTGATATTAGAATGAAA 89457

Qy 1422 TCTGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAAAT 1481
 ||| ||| ||||| ||| ||||| ||||| ||||| ||||| || || ||
 Db 89456 TCTAATGA--TGTCTGACTGTGAATAGAACCTGCTACCAATGTGAAATCTATAGAAAGAT 89399

Qy 1482 CCCTGGAAAGAGTGGAAAGTCCTGCCTAACACGTAAGTGCCTTCTT 1527
 ||||| ||||| ||| ||||| ||||| ||| || ||| ||
 Db 89398 -CCTGGAAAGAGTATAAAATCCTGCCTAACATGTACATGAATTTCAT 89354

RESULT 13

AY195873

LOCUS AY195873 2351 bp mRNA linear ROD 01-JUN-2003

DEFINITION Mus musculus strain PERA/Ei ATP-binding cassette sub-family G member 5 (Abcg5) mRNA, complete cds.

ACCESSION AY195873

VERSION AY195873.1 GI:31322257

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2351)

AUTHORS Wittenburg,H., Lyons,M.A., Li,R., Churchill,G.A., Carey,M.C. and Paigen,B.

TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred Mice

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2351)

AUTHORS Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.

TITLE Direct Submission

JOURNAL Submitted (11-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

FEATURES Location/Qualifiers

source 1..2351

/organism="Mus musculus"

/mol_type="mRNA"

```

/strain="PERA/Ei"
/db_xref="taxon:10090"
/chromosome="17"
/map="55 cM"
/sex="male"
/tissue_type="liver"
gene      1. .2351
          /gene="Abcg5"
CDS       139. .2097
          /gene="Abcg5"
          /note="ATP-dependent canalicular cholesterol transporter;
          white subfamily"
          /codon_start=1
          /product="ATP-binding cassette sub-family G member 5"
          /protein_id="AAO45094.1"
          /db_xref="GI:31322258"
          /translation="MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHV
          SYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISG
          RLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS
          SADFYNNKVEAVMTELSLSHVADQMIGSYNFGGSISSGERRRVISIAQQLLDQPKVMMLD
          EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG
          TPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESD
          IYHKILENIERARYLKTLPVPFKTKDPPGMFGKLGVLRLRVTRNLMRNKQAVIMRLV
          QNLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVS
          DQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALL
          APHLIGEFLTLVLLGIVQNPINVSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
          TFQKYCCILVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRTAN
          FLILYGFIPALVILGIVIFKVRDYLISR"

```

ORIGIN

```

Query Match      18.1%; Score 284.4; DB 10; Length 2351;
Best Local Similarity 98.0%; Pred. No. 1.1e-76;
Matches 288; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy      285 ATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCC 344
          |||
Db        1 ATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCC 60

Qy      345 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCATTGTC 404
          |||
Db        61 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCATTGTC 120

Qy      405 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 464
          |||
Db       121 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 180

Qy      465 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 524
          |||
Db       181 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 240

Qy      525 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTTACAGCGTCAGGTAAGGGGAC 578
          |||
Db       241 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTTACAGCGTCAGCAACCGTGTC 294

```

RESULT 14
AX456524

LOCUS AX456524 2354 bp DNA linear PAT 06-JUL-2002
 DEFINITION Sequence 46 from Patent WO0227016.
 ACCESSION AX456524
 VERSION AX456524.1 GI:21715413
 KEYWORDS .
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.
 REFERENCE 1
 AUTHORS Patel,S.B. and Dean,M.
 TITLE Gene involved in dietary sterol absorption and excretion and uses therefor
 JOURNAL Patent: WO 0227016-A 46 04-APR-2002;
 THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
 Shailendra B. (US) ; Dean, Michael (US)
 FEATURES Location/Qualifiers
 source 1. .2354
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Primer"

ORIGIN

Query Match 18.1%; Score 284.4; DB 6; Length 2354;
 Best Local Similarity 98.0%; Pred. No. 1.1e-76;
 Matches 288; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 285 ATTGGTGAAGTGTATCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCC 344
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ATTGGTGAAGTGTATCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCC 60
 Qy 345 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGC 404
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGC 120
 Qy 405 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 464
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 180
 Qy 465 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 524
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 240
 Qy 525 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGAC 578
 |||||||||||||||||||||||||||||||||||||||| | | | |
 Db 241 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTGTC 294

RESULT 15

AF312713

LOCUS AF312713 2354 bp mRNA linear ROD 16-MAY-2001
 DEFINITION Mus musculus sterolin (Abcg5) mRNA, complete cds.
 ACCESSION AF312713
 VERSION AF312713.2 GI:14091944
 KEYWORDS .
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2354)

AUTHORS Lee,M.H., Lu,K., Hazard,S., Yu,H., Shulenin,S., Hidaka,H., Kojima,H., Allikmets,R., Sakuma,N., Pegoraro,R., Srivastava,A.K., Salen,G., Dean,M. and Patel,S.B.

TITLE Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption

JOURNAL Nat. Genet. 27 (1), 79-83 (2001)

MEDLINE 20578753

PUBMED 11138003

REFERENCE 2 (bases 1 to 2354)

AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

REFERENCE 3 (bases 1 to 2354)

AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

REMARK Sequence update by submitter

COMMENT On May 16, 2001 this sequence version replaced gi:12382299.

FEATURES

Location/Qualifiers

source 1. .2354

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/tissue_type="liver"

gene 1. .2354

/gene="Abcg5"

CDS 139. .2097

/gene="Abcg5"

/note="ABCG5"

/codon_start=1

/product="sterolin"

/protein_id="AAG53097.1"

/db_xref="GI:12382300"

/translation="MGELPFLSPEGARGPHINRGSLSLSLEQGSVTGTEARHSLGVLHV
SYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISG
RLRRTGTLEGEVFVNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLYTAMLALCRS
SADFYNNKKVEAVMTLSLSHVADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLD
EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG
TPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESD
IYHKILENIERARYLKTLPVTPFKTKDPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLV
QNLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMNAVNLFPMLRAVS
DQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALL
APHLIGEFLTLVLLGIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
TFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRTAN
FLILYGFIPALVILGIVIFKVRDYILISR"

ORIGIN

Query Match 18.1%; Score 284.4; DB 10; Length 2354;
Best Local Similarity 98.0%; Pred. No. 1.1e-76;

Matches 288; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Qy      285 ATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCC 344
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  ATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCC 60

Qy      345 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGC 404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGC 120

Qy      405 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 180

Qy      465 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAG 524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAG 240

Qy      525 GCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTACAGCGTCAGGTAAGGGGAC 578
          |||||||||||||||||||||||||||||||||||||||| | | | |
Db      241 GCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTACAGCGTCAGCAACCGTGTG 294

```

RESULT 16

AY195872

LOCUS AY195872 2351 bp mRNA linear ROD 01-JUN-2003

DEFINITION Mus musculus strain I/LnJ ATP-binding cassette sub-family G member 5 (Abcg5) mRNA, complete cds.

ACCESSION AY195872

VERSION AY195872.1 GI:31322255

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2351)

AUTHORS Wittenburg,H., Lyons,M.A., Li,R., Churchill,G.A., Carey,M.C. and Paigen,B.

TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred Mice

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2351)

AUTHORS Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.

TITLE Direct Submission

JOURNAL Submitted (11-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

FEATURES Location/Qualifiers

```

source      1. .2351
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="I/LnJ"
              /db_xref="taxon:10090"
              /chromosome="17"
              /map="55 cM"
              /sex="male"
              /tissue_type="liver"

```

gene 1. .2351
 /gene="Abcg5"
 CDS 139. .2097
 /gene="Abcg5"
 /note="ATP-dependent canalicular cholesterol transporter;
 white subfamily"
 /codon_start=1
 /product="ATP-binding cassette sub-family G member 5"
 /protein_id="AAO45093.1"
 /db_xref="GI:31322256"
 /translation="MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHV
 SYSVSNRVGPWWNIKSCQQKWDRIKLDVSLYIESGQIMCILGSSSGSKTTLDDAISG
 RLRCCTGTLEGDFVNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS
 SADFYNNKVEAVMTELSLSHVADQVIGSYNFGGIISSGERRRVSIAAQLLQDPKVMMLD
 EPTTGLDCMTANQIVLLLAELARRDRIVITIHQPRSELFQHFQDKIAILTYGELVFCG
 TPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLESFAKESD
 IYHKILENIERARYLKTLPVFPKTKDPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLV
 QNLIMGLFLIFYLLRVQNNLTGKAVQDRVGLLYQFVGATPYTGMNAVNLFPMRLAVS
 DQESQDGLYHKWQMLLAYVLHALPFSIIATVIFSSVCYWTGLGYPEVARFGYFSAALL
 APHLIGEFLTLVLLGIVQNPINVSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
 TFQKYCCEILVNEFYGLNFTCGESNTTMLNHPMAITQGVFIEKTCPGATSRTAN
 FLILYGFIPALVILGIVIFKVRDYLISR"

ORIGIN

Query Match 18.0%; Score 282.8; DB 10; Length 2351;
 Best Local Similarity 97.6%; Pred. No. 3.4e-76;
 Matches 287; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy	285	ATTGGTGAAGTGTATCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCC	344
Db	1	ATTGGTGAAGTGTATCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCC	60
Qy	345	CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGC	404
Db	61	CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGC	120
Qy	405	ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG	464
Db	121	ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG	180
Qy	465	CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG	524
Db	181	CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACGGAG	240
Qy	525	GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGAC	578
Db	241	GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTCT	294

RESULT 17

AC084265/c

LOCUS AC084265 127066 bp DNA linear PRI 11-DEC-2001
 DEFINITION Homo sapiens chromosome 2, clone CTB-2367F13, complete sequence.
 ACCESSION AC084265
 VERSION AC084265.4 GI:17488659
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 2, clone CTB-2367F13

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnéz,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (18-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (24-AUG-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Dec 11, 2001 this sequence version replaced gi:15284200.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11578
Center clone name: 2367_F_13

FEATURES Location/Qualifiers

source 1. .127066
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="CTB-2367F13"
/clone_lib="CITB Human BAC"

repeat_region complement(8. .170)
/rpt_family="MER47A"

repeat_region 171. .468
/rpt_family="AluSx"

repeat_region complement(469. .516)
/rpt_family="MER47A"

repeat_region 988. .1049
/rpt_family="MIR"

repeat_region complement(1294. .1448)

repeat_region	/rpt_family="L1ME4A"
	complement(2662. .2954)
repeat_region	/rpt_family="AluSx"
	4049. .4431
unsure	/rpt_family="L2"
	5261. .5269
unsure	/note="<30 qual SNGL region"
	7192. .7202
repeat_region	/note="<30 qual SNGL region"
	7310. .7472
repeat_region	/rpt_family="MIR"
	7488. .7582
repeat_region	/rpt_family="MIR"
	7589. .7628
repeat_region	/rpt_family="(TTG)n"
	complement(7631. .7781)
repeat_region	/rpt_family="AluSg/x"
	7791. .7922
repeat_region	/rpt_family="MIR"
	complement(7977. .8300)
repeat_region	/rpt_family="AluJb"
	9044. .9343
repeat_region	/rpt_family="AluSq"
	10315. .10344
repeat_region	/rpt_family="AT_rich"
	10355. .10681
repeat_region	/rpt_family="AluJo"
	10683. .10993
repeat_region	/rpt_family="AluSx"
	complement(12221. .12282)
repeat_region	/rpt_family="MIR3"
	complement(12306. .12449)
repeat_region	/rpt_family="MIR"
	complement(13008. .13189)
repeat_region	/rpt_family="MER33"
	complement(13190. .13471)
repeat_region	/rpt_family="AluJo"
	complement(13472. .13612)
repeat_region	/rpt_family="MER33"
	13899. .13922
repeat_region	/rpt_family="GC_rich"
	complement(14184. .14250)
repeat_region	/rpt_family="L2"
	14552. .14630
repeat_region	/rpt_family="MER5A"
	complement(14809. .15100)
repeat_region	/rpt_family="AluSx"
	complement(15363. .15679)
repeat_region	/rpt_family="AluY"
	complement(15681. .15979)
repeat_region	/rpt_family="AluSx"
	16292. .16388
repeat_region	/rpt_family="L2"
	16392. .16508
repeat_region	/rpt_family="MLT1F"
	complement(16538. .16616)
repeat_region	/rpt_family="LTR37B"

```

repeat_region 16618. .16687
                /rpt_family="Alu"
repeat_region complement(16988. .17104)
                /rpt_family="L2"
repeat_region 17540. .17895
                /rpt_family="MLT1A1"
repeat_region complement(17911. .18209)
                /rpt_family="AluSq"
repeat_region 18487. .18680
                /rpt_family="LTR16A1"
repeat_region 18802. .19026
                /rpt_family="AluJo"
repeat_region complement(19092. .19390)
                /rpt_family="AluJo"
repeat_region complement(21369. .21675)
                /rpt_family="AluSx"
repeat_region complement(22474. .22763)
                /rpt_family="MER115"
repeat_region complement(22843. .22942)
                /rpt_family="MER115"
repeat_region 23239. .23311
                /rpt_family="L3"
repeat_region complement(23968. .24265)

```

```

Query Match          17.7%;  Score 278.6;  DB 9;  Length 127066;
Best Local Similarity 55.7%;  Pred. No. 8.1e-75;
Matches 788;  Conservative 2;  Mismatches 561;  Indels 64;  Gaps 11;

```

```

Qy      120 CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
          || || | | | |||| ||| ||||| ||| | | | | || || |||||
Db      20948 CTCTGTTTCCTGGAGCAGGGACACCTCGGCCCTGCCCTGGGCCCGTCTCTCCCAGCAT 20889

Qy      180 TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTGCGAGATAAG 239
          ||||: ||||| | | | | | | | | | | | | | | | | | | | | |
Db      20888 TCCTTGCTGGCAAGCCCACCTACAAACGT----GTGTGTTCTGCCCAGTGTCAAGATAAG 20833

Qy      240 GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATGGTGAACGTGTTA 299
          ||| | ||||| ||||| ||||| ||||| || ||||| | | ||| |
Db      20832 GACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAAGTCCCAGTCTGCTG 20773

Qy      300 TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGT 359
          || || | | | | | | | | | | | | | | | | | | | | |
Db      20772 TCCCAAGGGACTCCGGGGTCAAGTGGAGCAGGCAGGGCAGTCTGCCACGGGGTCCCCAAC 20713

Qy      360 TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTGCATTTGCTTCCTGCTA 419
          | |||| | |||| | || ||||| ||||| | | ||||| |||| |
Db      20712 TGAAGCCACTCTGGGGAGGGTCCGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTG 20653

Qy      420 GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA 479
          ||||| ||||| || | || | || | || | || | || | |||||
Db      20652 GCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA 20593

Qy      480 GGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTA 539
          || || | ||||| ||||| | | | | || | | | || | |||||
Db      20592 GGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTG 20536

Qy      540 GGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGC 599

```

Db	20535	GGCATCCTCCATGCCTCCTACAGCGTCAGGTAAGGCAGAGCCCTT-----GCTGCTG	20484
Qy	600	TCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTGTCTGTCCAG	659
Db	20483	CTGCTCCCCCAGGAGTGCAGGGGCGGCGCTCACCCCTCTGCTGCCTTTCTTCACTCTTT	20424
Qy	660	CAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCT	719
Db	20423	AAGTGCCAGTCTGGGCACTTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGGAGAGG	20364
Qy	720	GCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTG-CTTCTC	778
Db	20363	GAGAAGGGCTGTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCCTCCCCCTGCCCAGGG	20304
Qy	779	ACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCGTGTGTAGATGGAGAA	838
Db	20303	AAGGGCCTGGGCATAAACAAAGTGGCAGCAGTGCCCTGCCAACCCAGTGTCTACGGCCTG	20244
Qy	839	GGCTCG-----GAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGG	892
Db	20243	CCCTCTGTGGATGGGAATGGGGGTACTGCGAATGCAAGGAGTCTTGAAACCTGGTGAAAG	20184
Qy	893	AATGCAGGGTTCACCTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGA	952
Db	20183	AATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTTCAGAGCAACTCC	20130
Qy	953	CCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTCATT	1012
Db	20129	AGCACAGGCCCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTTCTACAG	20070
Qy	1013	ATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAG	1072
Db	20069	AGGAGGGCG-----CAGAGACTGAAACACGTTAGGAGCCTGTCCGGAGACTACTGGGG	20017
Qy	1073	GGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTC	1132
Db	20016	TGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCCTTCCAGGGCCCCAAGC	19957
Qy	1133	TGCCTTTGCCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGCCTTGGTGGAACATCA	1192
Db	19956	TGCCTTTGCCCTTCTTGGGGTTTCTTTAAAGCCACCGCGTGAGGCCCTGGTGGGACATCA	19897
Qy	1193	AATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGA	1252
Db	19896	CATCTTGCCGGCAGCAGTGGACCAGGACAGATCCTCAAAGATGTCTCCTTGTACGTGGAGA	19837
Qy	1253	GTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCC	1312
Db	19836	GCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTAAG-----	19799
Qy	1313	TGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAATGTTTAAGAAG	1372
Db	19798	---CTTGGGAAGGAGGATTCTAAAAAGGATTGGCTTGAGTTAAACTCCACATTGAAGAA	19742
Qy	1373	TRAGTTTAAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATCTGGTGACTT	1432

Db 19741 ACAGATTAAGTTGTAACAAGAAAGCCACAGGTTTGATATTAGAATGAATTCTATTGA--T 19684

Qy 1433 GTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGA 1492
 |||| ||| ||| ||| || ||||| ||||| || || || ||| |||||

Db 19683 ATCTGACTGTGAATGGAA-CTGCTACCAATGTGAAATCTTTAGAAAGAT-CCTTGAAAGA 19626

Qy 1493 GTGGAAAGTCCTGCCTAACACGTAAGTGCCTTCTT 1527
 || ||| | ||||| ||| ||| ||| |

Db 19625 GTATAAAATTCTGCCTAACATGTACGTGAATTCAT 19591

RESULT 18

AC108476/c

LOCUS AC108476 139342 bp DNA linear PRI 16-APR-2002

DEFINITION Homo sapiens BAC clone RP11-1413K20 from 2, complete sequence.

ACCESSION AC108476

VERSION AC108476.5 GI:19807988

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 139342)

AUTHORS Sulston, J.E. and Waterston, R.

TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792

PUBMED 9847074

REFERENCE 2 (bases 1 to 139342)

AUTHORS Harkins, C., Haakenson, W. and Doebber, A.

TITLE The sequence of Homo sapiens BAC clone RP11-1413K20

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 139342)

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (27-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 4 (bases 1 to 139342)

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (20-FEB-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 5 (bases 1 to 139342)

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 6 (bases 1 to 139342)

AUTHORS Waterston, R.

TITLE Direct Submission

JOURNAL Submitted (16-APR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT On Mar 29, 2002 this sequence version replaced gi:18767626.

----- Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

----- Summary Statistics

Center project name: H_NH1413K20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenos, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-489K22, 2000 bp overlap. Actual end is at base position 139342 of RP11-1413K20.

The region between 132012 and 132017 is covered only by a pcr product of clone DNA.

FEATURES	Location/Qualifiers
source	1. .139342 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="2" /map="2" /clone="RP11-1413K20" /clone_lib="RPCI-11"
misc_feature	55. .655 /note="match to EST AA203458 (NID:g1799169) zx58b04.r1"
misc_feature	93. .286 /note="match to EST AV689089 (NID:g10290952)"
misc_feature	93. .286

```

                                /note="similar to Mus musculus EST AI597378 (NID:g4606426)
                                vj29c06.y1"
misc_feature      93. .279
                                /note="match to EST AV660973 (NID:g9881987)"
misc_feature      318. .653
                                /note="match to EST R00405 (NID:g750141) ye71e05.r1"
misc_feature      372. .633
                                /note="similar to Homo sapiens EST T97887 (NID:g747232)
                                ye58h05.r1"
misc_feature      706. .708
                                /note="match to EST R00405 (NID:g750141) ye71e05.r1"
misc_feature      706. .707
                                /note="similar to Homo sapiens EST T97887 (NID:g747232)
                                ye58h05.r1"
repeat_region     847. .1139
                                /rpt_family="Alu"
misc_feature      1867. .2047
                                /note="match to EST T39945 (NID:g647612) ya13g04.r1"
repeat_region     2234. .2616
                                /rpt_family="L2"
misc_feature      2983. .3121
                                /note="match to EST AV689089 (NID:g10290952)"
misc_feature      2983. .3121
                                /note="similar to Mus musculus EST AI597378 (NID:g4606426)
                                vj29c06.y1"
misc_feature      3044. .3121
                                /note="match to EST T86384 (NID:g714736) yd77b08.r1"
misc_feature      4099. .4304
                                /note="match to EST T86384 (NID:g714736) yd77b08.r1"
misc_feature      4099. .4283
                                /note="match to EST AV689089 (NID:g10290952)"
misc_feature      4401. .4618
                                /note="similar to Mus musculus EST BF162656
                                (NID:g11042879)"
misc_feature      4405. .4454
                                /note="match to EST T86384 (NID:g714736) yd77b08.r1"
misc_feature      4724. .5110
                                /note="similar to Homo sapiens EST AV656623
                                (NID:g9877637)"
misc_feature      5075. .5204
                                /note="similar to Mus musculus EST BF162656
                                (NID:g11042879)"
repeat_region     5495. .5657
                                /rpt_family="MIR"
repeat_region     5673. .5767
                                /rpt_family="MIR"
repeat_region     5774. .5813
                                /rpt_family="(TTG)n"
repeat_region     5816. .5958
                                /rpt_family="Alu"
repeat_region     5976. .6091
                                /rpt_family="MIR"
repeat_region     6162. .6485
                                /rpt_family="Alu"
misc_feature      6351. .6373
                                /note="match to EST AA228345 (NID:g1849916) nc39d04.s1"
misc_feature      6352. .6364

```

```

misc_feature      /note="match to EST AI431309 (NID:g4302284) ar55b01.xl"
6352. .6364
misc_feature      /note="match to EST AI469772 (NID:g4331862) tm20f11.xl"
6353. .6367
misc_feature      /note="match to EST AI241685 (NID:g3837082) qu70f06.xl"
6568. .6707
misc_feature      /note="similar to Mus musculus EST BF162656
(NID:gl1042879)"
6649. .6707
misc_feature      /note="similar to Mus musculus EST BB598373
(NID:gl6450340)"
7229. .7528
repeat_region    /rpt_family="Alu"
7940. .8549
misc_feature      /note="similar to EST BM725726 (NID:g19047059)"
8169. .8305
misc_feature      /note="similar to Mus musculus EST BF162656
(NID:gl1042879)"
8169. .8301
misc_feature      /note="similar to Mus musculus EST BB598373
(NID:gl6450340)"
8500. .8529
repeat_region    /rpt_family="AT_rich"
8540. .8868
repeat_region    /rpt_family="Alu"
8870. .9180
repeat_region    /rpt_family="Alu"
10493. .10636
repeat_region    /rpt_family="MIR"
11195. .11376
repeat_region    /rpt_family="MER1_type"
11377. .11658
repeat_region    /rpt_family="Alu"
11659. .11799
repeat_region    /rpt_family="MER1_type"
11955. .12053
misc_feature      /note="similar to Mus musculus EST BB598373
(NID:gl6450340)"
11994. .12053
misc_feature      /note="similar to Mus musculus EST AA239884 (NID:gl863923)
mx81d01.r1"
12086. .12109
repeat_region

```

```

Query Match      17.7%; Score 278.6; DB 9; Length 139342;
Best Local Similarity 55.7%; Pred. No. 8.2e-75;
Matches 788; Conservative 2; Mismatches 561; Indels 64; Gaps 11;

```

```

Qy      120 CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
      || || | | | ||| ||| ||| || | | | | | | | | | |
Db      19164 CTCTGTTTCCTGGAGCAGGGACACCTCGGCCTCCTGCCCTGGGCCCGTCTCTCCCAGCAT 19105

Qy      180 TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTGCGAGATAAG 239
      ||||: ||||| | | | | | | | | | | | | | | | |
Db      19104 TCCTTGCTGGCAAGCCACCTACAAACGT----GTGTGTTCTGCCCACTGTCAAGATAAG 19049

Qy      240 GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTTA 299
      ||| | ||||| ||||| ||||| || | ||||| | | | |

```

Db	19048	GACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCAGTCTCTGCTG	18989
Qy	300	TCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT	359
Db	18988	TCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAAC	18929
Qy	360	TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCATTTGCTTCCTGCTA	419
Db	18928	TGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTG	18869
Qy	420	GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA	479
Db	18868	GCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA	18809
Qy	480	GGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTA	539
Db	18808	GGCTCCACAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTG	18752
Qy	540	GGTGTCTTGCATGTGTCTTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGC	599
Db	18751	GGCATCTCTCCATGCCTCTTACAGCGTCAGGTAAGGCAGAGCCCTT-----GCTGCTG	18700
Qy	600	TCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTGTCTGTCCAG	659
Db	18699	CTGCTCCCCCAGGAGTGCGGGGGCCCGGCGCTACCCCTCTGCTGCCTTTCTTCACTCTTT	18640
Qy	660	CAGATCAGGGTGAAAGTGGACAGTCTGTAAACAACAGTGAGTCGTTCCCTCCTCCTCCT	719
Db	18639	AAGTGCCAGTCTGGGCACTTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGGAGAGG	18580
Qy	720	GCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTG-CTTCTC	778
Db	18579	GAGAAGGGCTGTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCCTCCCCCTGCCAGGG	18520
Qy	779	ACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCTGTGTAGATGGAGAA	838
Db	18519	AAGGGCCTGGGCATAAAACAAAGTGGCAGCAGTGCCCTGCCAACCCAGTGTCTACGGCCTG	18460
Qy	839	GGCTCG-----GAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGG	892
Db	18459	CCCTCTGTGGATGGGAATGGGGGTACTGCGAATGCAAGGAGTCTTGAAACCTGGTGAAAG	18400
Qy	893	AATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGA	952
Db	18399	AATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTCAGAGCAACTCC	18346
Qy	953	CCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTCAAT	1012
Db	18345	AGCACAGGCCCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTTCTACAG	18286
Qy	1013	ATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAG	1072
Db	18285	AGGAGGGCG-----CAGAGACTGAAACACGTTAGGAGCCTGTCCGGAGACTACTGGGG	18233
Qy	1073	GGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTC	1132
Db	18232	TGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCCCTTCCAGGGCCCCAAGC	18173

Qy 1133 TGCCTTTGCCCTTGTGGGATTTCTTTAAAGCAACCGTGTCCGGGCCCTTGGTGGAAACATCA 1192
 ||||| | |||| | ||||| |||| | |||| | ||||
 Db 18172 TGCCTTTGCCCTTCTGGGGTTTCTTTAAAGCCACCGCGTGAGGCCCTGGTGGGACATCA 18113

Qy 1193 AATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGA 1252
 ||| |||| |||| |||| | |||| | ||||| ||||| ||||| ||||| ||||| |||||
 Db 18112 CATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGA 18053

Qy 1253 GTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCC 1312
 | || |||| | ||||| |||| | ||||| |||||
 Db 18052 GCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTAAG----- 18015

Qy 1313 TGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAG 1372
 | |||| | ||| | | | | | : | | | | || |
 Db 18014 ---CTTGGGAAGGAGGATTCTAAAAAGGATTTGGCTTGAGTTAAACTCCACATTGAAGAA 17958

Qy 1373 TRAGTTTAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATCTGGTGACTT 1432
 || ||||| || || |||| | || | || | || | || |
 Db 17957 ACAGATTAAGTTGTAACAAGAAAGCCACAGGTTTGATATTAGAATGAATCTATTGA--T 17900

Qy 1433 GTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGA 1492
 |||| || | || | || | |||| | ||||| || || || || || |||||
 Db 17899 ATCTGACTGTGAATGGAA-CTGCTACCAATGTGAAATCTTTAGAAAGAT-CCTTGAAAGA 17842

Qy 1493 GTGGAAAGTCCTGCCTAACACGTAAGTGCCTTCTT 1527
 || || | ||||| || || || || |
 Db 17841 GTATAAAATTCTGCCTAACATGTACGTGAATTCAT 17807

RESULT 19

AC145533

LOCUS AC145533 159346 bp DNA linear HTG 19-JUL-2003

DEFINITION Lemur catta clone LB2-138H20, WORKING DRAFT SEQUENCE, 5 unordered pieces.

ACCESSION AC145533

VERSION AC145533.1 GI:32996774

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Lemur catta (ring-tailed lemur)

ORGANISM Lemur catta

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Strepsirhini; Lemuridae; Lemur.

REFERENCE 1 (bases 1 to 159346)

AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R., Peng,Z., Malinov,I. and Rubin,E.M.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 159346)

AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R., Peng,Z., Malinov,I. and Rubin,E.M.

TITLE Direct Submission

JOURNAL Submitted (19-JUL-2003) Genome Sciences, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

COMMENT Draft Sequence Produced by Berkeley PGA

Web site: <http://pga.lbl.gov>

Center Code: PGABERK

Center Project Name: L105-138H20

Bac Clone Name: LB2-138H20

Additional information on comparative analysis and ordering are available at:

http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=ABCG5

Funding agent: Programs for Genomic Applications (NHLBI)

if library name is LB1 to LB4, please see website

for a description: <http://www-gsd.lbl.gov/cheng/BAC.html>

These libraries are available through the BACPAC Resources Center:

<http://www.chori.org/bacpac/libraryres.htm> as LBNL-1 to LBNL-4.

Summary Statistics:

Sequencing vector: Plasmid; pUC18

Chemistry: Dye-terminator Big Dye

Assembly program: Phrap version 0.990329.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 16021: contig of 16021 bp in length

* 16022 16121: gap of unknown length

* 16122 40145: contig of 24024 bp in length

* 40146 40245: gap of unknown length

* 40246 77537: contig of 37292 bp in length

* 77538 77637: gap of unknown length

* 77638 114811: contig of 37174 bp in length

* 114812 114911: gap of unknown length

* 114912 159346: contig of 44435 bp in length.

FEATURES

source

Location/Qualifiers

1. .159346

/organism="Lemur catta"

/mol_type="genomic DNA"

/db_xref="taxon:9447"

/clone="LB2-138H20"

ORIGIN

Query Match 17.5%; Score 275.4; DB 2; Length 159346;

Best Local Similarity 58.2%; Pred. No. 8.2e-74;

Matches 781; Conservative 2; Mismatches 458; Indels 102; Gaps 13;

```
Qy      230 TCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGG 289
          ||||| || || ||||| ||||| ||||| ||||| |||||
Db      89750 TCGAGATAAGAACTGTCCGGCTAAAGGTTTCATCAGATAATGGCATCTGTGGCCAAACCCC 89809

Qy      290 TGAACGTGT-----ATCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGG 338
          || || | || | | | | | | | | | | | | | | |
Db      89810 TGCTCTGCCTCCAGAACGGCATCACGAGGACTCCAGGCCAGGAGAGGAGCAGGCAGGG 89869

Qy      339 CA--CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAAT 396
          || || | ||||| | || | | | | | | | | | |
Db      89870 TCACCTGCCACCAGCTCCTCAGCTGAAGCCACTCCGGGGAGCGACAGGTGGCCAGAAAAC 89929

Qy      397 TCACTTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAG 456
          || | | | |||| | || | | | | | | | | | |
```

Db	89930	TCACCCGTCTCTGCTGCCTGCTGGCCATGAGCGACCTCCTATCGTTGGCGCCAGGGGGT	89989
Qy	457	CCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGG	516
Db	89990	CCTTGGGCCTCCCCGTGAGCAGAGGCCCGGAGCTCTCTGGAGGAGGTTCTGCCGCCG	90049
Qy	517	GCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCCTACAGCGTCAGGTAAG---	573
Db	90050	CTTCGAGCCGCGGCACTGCCTGGGCATCTCCCATGCCTCCTACAGCGTCAGGTAAGCCA	90109
Qy	574	-----GGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGGG	628
Db	90110	CAGCCTCCCCCTGCCGGCCTGAGGCCTGGGCTCTCACCCCTCTGCTCACTCATGGGGGC	90169
Qy	629	TGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATC-----AGGGTGAAAGT	676
Db	90170	CCTGCCTGGGGCTTTTCAGGCTCCCTCTTCAGTGGCCCTCAAGGGGGGAAGAGGAGGAAGC	90229
Qy	677	GGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTG	736
Db	90230	GCCTGGGAAATGGGGAGCAACAGTGAACGCCCTCCTCCTGCCAGGCCAGGCCCGGCCCA	90289
Qy	737	GACATTAAAAATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCC	796
Db	90290	GGCACAAGCAATGTTGCCAACCTGCGCCCTGACTCCTACCCTCCCCATTGCCACCCCT	90349
Qy	797	CTTCCTTGACTCGCCCACCACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGT	856
Db	90350	GTGTCTGTGG-----CTGCCCTGTGTGGATGCCGG-GACCAGCGGAATGGGGGT	90397
Qy	857	GCTGGGGGACAAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACTTCAAGAAGA	916
Db	90398	GCTGGGAACATGAGGGGTAATAAAGCCTGGGGAAGCAATGCAGGGTCAACAAC-----	90450
Qy	917	AAGCAGTGTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCT-AATG	975
Db	90451	-----CTTCCGGTCAGACCGAGAGGGGACATTCAGAACAGCTGCAGG	90492
Qy	976	GGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTCAATTATACCTCCAAGGACAACAGAGTG	1035
Db	90493	CCAGGCCCTCTCCGCAGGTGATGGGCAGCCTCGCCACTGCCCGCCAGGCTCTGCAGAGGA	90552
Qy	1036	GTACATAAGGCTAAAAACAGATTGTCAACCTGTCCAGGGGCAACTGGGATGGG-----	1088
Db	90553	GGGCACAGAGGCTGAATCACCTTAGGAACCTGTCCAGAGACAACCTGGGGTGGGTGCAGCT	90612
Qy	1089	----GTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCT	1144
Db	90613	GGGATCAGAGCTGGGGGACGGGGTCTGGCCTGTTCCAGGCCCCCATGCTGTCTTTGCCCT	90672
Qy	1145	TGTGGGATTTCTTTAAAGCAACCGTGTCTGGGCCTTGGTGGAACATCAAATCATGCCAGC	1204
Db	90673	CCCGGGGTTTCTTTAAAGCAATCGTGTCTGGGCCCTGGTGGGACATTGCATCTTGCCGGA	90732
Qy	1205	AGAAGTGGGACAGGCAAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTA	1264
Db	90733	AGCAGTGGAAACAGGCAGATCCTCAAAGACGTCTCCTTGTACGTTGAGAGTGGGCAGATCA	90792

Qy 1265 TGTGCATCTTAGGCAGCTCAGGTAAGTGCC TGGGGGGSCSGGGGCTCCTGTACTTCTAAG 1324
 ||||| ||| ||||| ||| | ||: || | || | ||| ||
 Db 90793 TGTGCATTCTAGGGAGCTCAGGTAA-----GCTGGGAAGGAGTTCTCTGAGTTCTCAG 90845

Qy 1325 GCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTAAGTT 1384
 | ||| |||| | ||| ||| | | |||| : ||||| |||
 Db 90846 -----TGAAGGGTTTGGTTTGATCTA--CACCACAGTGAAGAAACAGGTTTAAGTT 90894

Qy 1385 GTAGAGAGGCAGCCATGCATTTGGCATTTGAATACAATCTGGTGACTTGTCTGGCTGCCA 1444
 | | || || ||| || ||| ||| ||| ||| ||||| ||| |
 Db 90895 GCTGCAAGAAGTCCGCAAGTTTGATATCAGAATGAAATTAAATGACATGTCTGACTGTGA 90954

Qy 1445 ATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCT--GGAAAGAGTGAAAGTC 1502
 || ||| || ||| ||| ||||| || || ||| | ||||| ||| |
 Db 90955 ATGGAATCTGGTATCAATGTGAAATCTTTAGAAAGATCTTAAAAAAGAGTATAAAATT 91014

Qy 1503 CTGCCTAACACGTAAGTGCCTTC 1525
 | |||| |||| |||
 Db 91015 CCACCTAATGTATAAGTGAATTC 91037

RESULT 20

AC146286/c

LOCUS AC146286 207760 bp DNA linear HTG 15-AUG-2003

DEFINITION Callicebus moloch clone LB5-414K16, WORKING DRAFT SEQUENCE, 2
 ordered pieces.

ACCESSION AC146286

VERSION AC146286.2 GI:33667134

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE Callicebus moloch (Dusky titi)

ORGANISM Callicebus moloch

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Callicebinae;
 Callicebus.

REFERENCE 1 (bases 1 to 207760)

AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
 Peng,Z., Malinov,I. and Rubin,E.M.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 207760)

AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
 Peng,Z., Malinov,I. and Rubin,E.M.

TITLE Direct Submission

JOURNAL Submitted (02-AUG-2003) Genome Sciences, Lawrence Berkeley National
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

REFERENCE 3 (bases 1 to 207760)

AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
 Peng,Z., Malinov,I. and Rubin,E.M.

TITLE Direct Submission

JOURNAL Submitted (15-AUG-2003) Genome Sciences, Lawrence Berkeley National
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

COMMENT On Aug 15, 2003 this sequence version replaced gi:33413351.

Sequence Produced by Berkeley PGA

Web site: <http://pga.lbl.gov>

Center Code: PGABERK

Center Project Name: T039
Bac Clone Name: LB5-414K16

This sequence has been compared to sequences of other species using Vista (<http://www-gsd.lbl.gov/VISTA>). The results can be viewed at:
http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=ABCG5

The order-orientation of the draft sequence was accomplished by using:
Avid (<http://baboon.math.berkeley.edu/mavid>),
Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

If the Bac Library Name is LB1 to LB4, please see website for the description: <http://www-gsd.lbl.gov/cheng/BAC.html>
These libraries are available through the BACPAC Resources Center: <http://www.chori.org/bacpac/libraryres.htm> as LBNL-1 to LBNL-4.

Summary Statistics:

Sequencing vector: Plasmid; pUC18

Chemistry: Dye-terminator Big Dye

Assembly program: Phrap version 0.990329.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 74764: contig of 74764 bp in length
* 74765 74864: gap of unknown length
* 74865 207760: contig of 132896 bp in length.

FEATURES Location/Qualifiers
 source 1. .207760
 /organism="Callicebus moloch"
 /mol_type="genomic DNA"
 /db_xref="taxon:9523"
 /clone="LB5-414K16"

ORIGIN

Query Match 16.7%; Score 261.8; DB 2; Length 207760;
Best Local Similarity 56.2%; Pred. No. 1.5e-69;
Matches 803; Conservative 2; Mismatches 554; Indels 70; Gaps 14;

Qy 120 CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
 || || | | | |||| ||| ||||| |||| | | || | ||||| |||||
Db 140179 CTCTGTTTCCTGGAGCAGGGACACCTCAGCCTCCTGCCCTGGGCCCGGCTCTCCCAGCAT
140120

Qy 180 TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTGCGAGATAAG 239
 ||||:||||||| | | |||| | ||||| | ||||| ||||| | || |||||
Db 140119 TCCTCTCTGGCAAGCCCA-CCTACAAACACAT-GTGTGTTCTGCCCTCTCTCAAGATAAG
140062

Qy 240 GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTTA 299
 ||| | ||||| ||||| ||||| ||| || ||||| | | ||| |
 Db 140061 GACGCGCTGGCTAAAGGTACATCAGATAA-GGCCTCCGTGGCCAAGTCCCAGTCCTGCCA
 140003

Qy 300 TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT 359
 || | | | | | | | | | | | | | | | | | | | |
 Db 140002 TCCTGAGGGACTCCGGGGTCAGGTGGAGCCGGCAGGGCAGTCTGCCACTGGCTCGCCAAC
 139943

Qy 360 TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTGCATTTGCTTCCTGCTA 419
 | ||| | | | | | | ||||| ||||| || ||||| ||| |
 Db 139942 TGCAGCCACTCCGAGGAGGGTCAGGCCGCCAGAAAATCTGCTCAGCTTTGCTGCCCGTTG
 139883

Qy 420 GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA 479
 ||||| ||||| || | | | | | | | | | | | | | | | |
 Db 139882 GCCATGGGTGACCTTCCATCTTTGACCCCGGAGGGTCCATGGG---ACTCCTAAACAGA
 139826

Qy 480 GGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTA 539
 || |||| ||||| ||||| | | | | | | | | | | | | |
 Db 139825 GGCTCTCAGAGCTCCCTGGAGGGGGCTCCTGCCACTGCACCTGAGCCT---CACAGTCTG
 139769

Qy 540 GGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGG-----GGACCTCCACAG 586
 || ||||| ||||| ||||| ||||| ||||| | | |||| |
 Db 139768 GGAGTCCTGCATGCCTCCTACAGCATCAGGTAAGGCAGAGCCCTTGCTGCAGCTCCTCCC
 139709

Qy 587 CAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGG 646
 || | | | | | | | | | | | | | | | | | | | |
 Db 139708 CAGGAGCACGGGGCCCTGCGTTCACCCTCTGCTGCCTTTTTTCACTCCTGAGCTTCCTGG
 139649

Qy 647 GTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCC 706
 | | | | | | | | | | | | | | | | | | | | |
 Db 139648 CTGGGGACTTTGGGCTCCCTCTTCAGTGGATCAGGTGGAGAGAAGAGAGGGGGGAGGGCT
 139589

Qy 707 TCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCGC 766
 | | | | | | | | | | | | | | | | | | | | |
 Db 139588 GCACTGGGAAATGGGGAGCAACAGTGAATGGCCCCCTCCCCCTACCCAGGAAGGGCCTGG
 139529

Qy 767 TTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCTGT 826
 | | | | | | | | | | | | | | | | | | | | |
 Db 139528 GTATAAACAAAGTGGCAGCTGTGCCCTGCCTACCCAGTGT---CTACCGCCTGCCCTCT
 139472

Qy 827 GTAGATGGAGAAGGCTCGGAGAGTGGGGTGCTGGGGGCACAAAATGGAATGAACAC--- 883
 || ||||| || || ||||| ||||| | |||| | |||| |
 Db 139471 GTGGATGGAGAGAATCTGGGGAATGGGGG-GCTGGGAGTACAAGGAGTCTTGAAACCAGG
 139413

Qy 884 TGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCC 943

Db 139412 TGACGAATGCAGGGACAGTCACCTCCCAGCCAAATGGGCAGGACATTTCGGAGGAGCTCCA 139353
 Qy 944 AGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAA 1003
 Db 139352 GCACAGGCCCGCTCCCTAGGTGACAGACAGCCTCGGTCACTACCTGCCAGGTTCTACAGA 139293
 Qy 1004 CTTGTTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAA 1063
 Db 139292 GGAGGATGCCGAGGCTGAAACACATTAGAAACCTGTCTGAAGATAACTGG----- 139243
 Qy 1064 CCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGG 1123
 Db 139242 ---GGCGGGGGGACACAGGTGGGATCAATGCTGGGGACCTGGGTGTAGCCCCTTCCAGG 139186
 Qy 1124 ACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAGCAACCGTGTGGGGCCTTGGT 1183
 Db 139185 GCCACACGCTGCCTTTGCCCTTCTGGGATTTCTTTAAAGCCACCGTGTGGGGCCCTGGT 139126
 Qy 1184 GGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGT 1243
 Db 139125 GGGACGTCACATCTTGCCAGCGGCAGTGGACCAGGCAGATCCTCAAAGACGTCTCCTTGT 139066
 Qy 1244 ACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSC 1303
 Db 139065 ACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTAAGCTTGGGAAGAAG- 139007
 Qy 1304 SGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATG 1363
 Db 139006 -----GATTTTAAAAAGGCTTTGGCTTGAGTTAAACTCCACC 138970
 Qy 1364 TTTAAGAAGTRAGTTTAAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATC 1423
 Db 138969 CTGAAGAA-ACAGATACAGTTGTAGCAAGAAAGCTGCAGGTCTGATATTAGAAATGAAATC 138911
 Qy 1424 TGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAAATCC 1483
 Db 138910 TAATGA--TGTCTGACTGTGAATAGAACCTGCTACCAATGTGAAATCTATAGAAAGAT-C 138854
 Qy 1484 CTGGAAAGAGTGGAAAGTCCTGCCTAACACGTAAGTGCCTTCTTTGCTT 1532
 Db 138853 CTGGAAAGAGTATGAAATCCTGCCTAACATGTACATGAATTCATTATTT 138805

RESULT 21

AF404106

LOCUS AF404106

4899 bp

DNA

linear

PRI 14-AUG-2001

DEFINITION Homo sapiens clone BAC 328I4 sterolin 2 (ABCG8) and sterolin 1 (ABCG5) genes, partial cds.
 ACCESSION AF404106
 VERSION AF404106.1 GI:15150315
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 4899)
 AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H., Ose,L., Stalenhoef,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E., Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and Patel,S.B.
 TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by ABCG5 and ABCG8, respectively
 JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
 MEDLINE 21344600
 PUBMED 11452359
 REFERENCE 2 (bases 1 to 4899)
 AUTHORS Lu,K., Lee,M. and Patel,S.B.
 TITLE Direct Submission
 JOURNAL Submitted (11-JUL-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403
 FEATURES
 Location/Qualifiers
 source 1..4899
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="BAC 328I4"
 gene complement(<1..>3668)
 /gene="ABCG8"
 mRNA complement(<3606..>3668)
 /gene="ABCG8"
 /product="sterolin 2"
 CDS complement(<3606..3668)
 /gene="ABCG8"
 /codon_start=1
 /product="sterolin 2"
 /protein_id="AAK85386.1"
 /db_xref="GI:15150316"
 /translation="MAGKAAEERGLPKGATPQDTS"
 exon complement(3606..>3668)
 /gene="ABCG8"
 /number=1
 misc_feature 3669..4043
 /note="contains 5'UTR and promoter regions for ABCG5 and ABCG8"
 gene <4044..>4899
 /gene="ABCG5"
 mRNA join(<4044..4186,4770..>4891)
 /gene="ABCG5"
 /product="sterolin 1"
 CDS join(4044..4186,4770..>4891)
 /gene="ABCG5"

```

/codon_start=1
/product="sterolin 1"
/protein_id="AAK85387.1"
/db_xref="GI:15150317"
/translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
YSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSS"
exon <4044. .4186
/gene="ABCG5"
/number=1
exon 4770. .4891
/gene="ABCG5"
/number=2

```

ORIGIN

```

Query Match          15.6%;  Score 244.8;  DB 9;  Length 4899;
Best Local Similarity 55.0%;  Pred. No. 2.5e-64;
Matches 651;  Conservative 1;  Mismatches 488;  Indels 44;  Gaps 7;

```

```

Qy      120 CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
      || || | ||| ||| |||| ||| | | | | | | | | | | | |
Db      3744 CTCTGTTTCTTGGAGCAGGGACACCTCGGCCTCCTGCCCTGGGCCCTCTCTCCCAGCAT 3803

Qy      180 TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTGCGAGATAAG 239
      ||||: ||||| | | | | | | | | | | | | | | | | | |
Db      3804 TCCTTGCTGGCAAGCCCACC---TACAAACGTGTGTGTTCTTGCCCACTGTCAAGATAAG 3860

Qy      240 GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTTA 299
      || | | ||||| ||||| ||||| || | ||| | | | | | | |
Db      3861 GACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCAGTCCTGCTG 3920

Qy      300 TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT 359
      || || | | | | | | | | | | | | | | | | | | | | |
Db      3921 TCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAAC 3980

Qy      360 TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC TTGCATTTGCTTCCTGCTA 419
      | |||| | |||| | || | ||||| ||||| | | | | | | |
Db      3981 TGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTG 4040

Qy      420 GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA 479
      ||||| |||| | | | | | | | | | | | | | | | | |
Db      4041 GCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA 4100

Qy      480 GGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGTACGGGCACAGAGGCTCGGCACAGCTTA 539
      || | | | ||||| ||||| | | | | | | | | | | | | |
Db      4101 GGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTG 4157

Qy      540 GGTGTCCTGCATGTGTCTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGC 599
      || |||| |||| | ||||| ||||| | | | | | | | | |
Db      4158 GGCATCCTCCATGCCTCCTACAGCGTCAGGTAAGG-----CAGAGCCCTTGC 4204

Qy      600 TCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTGTCTGTCCAG 659
      | | | | | | | | | | | | | | | | | | | | | |
Db      4205 TGCTGCTGCTCCCCAGGAGTGCGGGGCCGGCGCTCACCCTCTGCTGCCTTTCTTCAC 4264

Qy      660 CAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCT 719
      | | | | | | | | | | | | | | | | | | | | |
Db      4265 TCTTTAAGTGCCAGTCTGGGCACCTTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGG 4324

```

Qy 720 GCGCAGGGCAGAGCCTGGACATTTAAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCA 779
 | | | | | | | | | | | | | | | | | | | |
 Db 4325 AGAGGGAGAAGGGCTGTTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCCTCCCCCTGC 4384
 Qy 780 CTGATTTCTGCTCTCCCTTCTCTTACTCGCC-----CACCACCTGTCCTGTGTAGAT 832
 | | | | | | | | | | | | | | | | | | | |
 Db 4385 CCAGGGAAGGGCCTGGGCATAAAACAAAGTGGCAGCAGTGCCCTGCCAACCCAGTGTCTAC 4444
 Qy 833 GGAGAAGGCTC-----GGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGC 886
 | | | | | | | | | | | | | | | | | | | |
 Db 4445 GGCCTGCCCTCTGTGGATGGGAATGGGGGTACTGCGAATGCAAGGAGTCTTGAACCTGG 4504
 Qy 887 TGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGT 946
 | | | | | | | | | | | | | | | | | | | |
 Db 4505 TGAAAGAATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTGAGAGC 4558
 Qy 947 CAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTT 1006
 | | | | | | | | | | | | | | | | | | | |
 Db 4559 AACTCCAGCACAGGCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTT 4618
 Qy 1007 GTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCT 1066
 | | | | | | | | | | | | | | | | | | | |
 Db 4619 CTACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCCT-----GTCCGGAGACTAC 4672
 Qy 1067 GTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACC 1126
 | | | | | | | | | | | | | | | | | | | |
 Db 4673 TGGGGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCCCTTCCAGGGCC 4732
 Qy 1127 CTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAGCAACCGTGTCCGGCCTTGGTGGG 1186
 | | | | | | | | | | | | | | | | | | | |
 Db 4733 CCAAGCTGCCTTTGCCCTTCTTGGGGTTTCTTTAAAGCCACCGGTGAGGCCCTGGTGGG 4792
 Qy 1187 ACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACA 1246
 | | | | | | | | | | | | | | | | | | | |
 Db 4793 ACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACG 4852
 Qy 1247 TCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAG 1290
 | | | | | | | | | | | | | | | | | | | |
 Db 4853 TGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTAAG 4896

RESULT 22

AX456521

LOCUS AX456521 5459 bp DNA linear PAT 06-JUL-2002

DEFINITION Sequence 43 from Patent WO0227016.

ACCESSION AX456521

VERSION AX456521.1 GI:21715411

KEYWORDS .

SOURCE synthetic construct

ORGANISM synthetic construct
 artificial sequences.

REFERENCE 1

AUTHORS Patel,S.B. and Dean,M.

TITLE Gene involved in dietary sterol absorption and excretion and uses therefor

JOURNAL Patent: WO 0227016-A 43 04-APR-2002;

THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
Shailendra B. (US) ; Dean, Michael (US)

FEATURES Location/Qualifiers
source 1. .5459
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Primer"

ORIGIN

Query Match 15.5%; Score 242.8; DB 6; Length 5459;
Best Local Similarity 54.9%; Pred. No. 1.1e-63;
Matches 649; Conservative 1; Mismatches 488; Indels 44; Gaps 7;

```
Qy      120 CTTTGCCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
      || || | ||| |||| ||| |||| ||| | | | | | |||||
Db      4309 CTCTGTTTCTTGAGCAGGGACACCTCGGCCTCCTGCCCTGGGCCGTCTCTCCCAGCAT 4368

Qy      180 TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTTCGAGATAAG 239
      ||||: ||||| | | | | | | | | | | | | | | | | |
Db      4369 TCCTTGCTGGCAAGCCCACC---TACAAACGTGTGTGTTCTTGCCCACTGTCAAGATAAG 4425

Qy      240 GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTTA 299
      ||| | ||||| ||||| ||||| ||| ||||| | | | | |
Db      4426 GACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCAGTCCCTGCTG 4485

Qy      300 TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT 359
      || || | | | | | | | | | | | | | | | | | |
Db      4486 TCCCAAGGGACTCCGGGGTTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGGCTCCCCAAC 4545

Qy      360 TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACTTGCAATTTGCTTCCTGCTA 419
      | |||| | |||| | || | ||||| ||||| | | ||||| |||| |
Db      4546 TGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTG 4605

Qy      420 GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA 479
      ||||| |||| | | | | | | | | | | | | | | | | |
Db      4606 GCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA 4665

Qy      480 GGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTA 539
      || || | ||||| ||||| | | | ||| | | | ||| | | ||||| |
Db      4666 GGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTG 4722

Qy      540 GGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGC 599
      || |||| |||| ||||| ||||| ||||| ||||| | | | | |
Db      4723 GGCATCCTCCATGCCTCCTACAGCGTCAGGTAAGG-----CAGAGCCCTTGC 4769

Qy      600 TCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAG 659
      | || || || || || | || | || | || || || || || ||
Db      4770 TGCTGCTGCTCCCCAGGAGTGCGGGGCCCGGCGCTCACCCCTCTGCTGCCTTTCTTCAC 4829

Qy      660 CAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTTCCTCCTCCTCCTCCT 719
      | || | | ||| || | | | | | | | | | | | | | |
Db      4830 TCTTTAAGTGCCAGTCTGGGCACCTTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGG 4889

Qy      720 GCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCA 779
      | | || || | ||||| || | | | | | | | | | |
Db      4890 AGAGGGAGAAGGGCTGTTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCCTCCCCCTGC 4949
```


Qy 780 CTGATTCTGCTCTCCCTTCCTTGACTCGCC-----CACCACCTGTCCTGTGTAGAT 832
 | | | | | | | | | | | | | | | | | | | |
 Db 4950 CCAGGAAGGGCCTGGGCATAAAACAAAGTGGCAGCAGTGCCCTGCCAACCCAGTGTCTAC 5009

Qy 833 GGAGAAGGCTC-----GGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGC 886
 || | | | | | | | | | | | | | | | | | | | |
 Db 5010 GGCCTGCCCTCTGTGGATGGGAATGGGGTACTGCGAATGCAAGGAGTCTTGAAACCTGG 5069

Qy 887 TGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGT 946
 |||| | | | | | | | | | | | | | | | | | | | |
 Db 5070 TGAAAGAATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTAGAGC 5123

Qy 947 CAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTT 1006
 | | | | | | | | | | | | | | | | | | | | | |
 Db 5124 AACTCCAGCACAGGCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTT 5183

Qy 1007 GTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCT 1066
 | | | | | | | | | | | | | | | | | | | | | |
 Db 5184 CTACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCCT-----GTCCGGAGACTAC 5237

Qy 1067 GTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACC 1126
 | | | | | | | | | | | | | | | | | | | | | |
 Db 5238 TGGGGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCTTCCAGGGCC 5297

Qy 1127 CTACTCTGCCTTTGCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCCGGGCCCTTGGTGGA 1186
 | | | | | | | | | | | | | | | | | | | | | |
 Db 5298 CCAAGCTGCCTTTGCCCTTCCCTGGGGTTTCCTTTAAAGCCACCGGTGAGGCCCTGGTGGA 5357

Qy 1187 ACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACA 1246
 ||||| | | | | | | | | | | | | | | | | | | | | |
 Db 5358 ACATCACATCTTGCCGGCAGCAGTGGACCAGGACAGATCCTCAAAGATGTCTCCTTGTACG 5417

Qy 1247 TCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTA 1288
 | | | | | | | | | | | | | | | | | | | | | |
 Db 5418 TGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTA 5459

RESULT 23

F351812S01/c

LOCUS F351812S01 2809 bp DNA linear PRI 10-AUG-2001

DEFINITION Homo sapiens sterolin-2 (ABCG8) gene, exon 1.

ACCESSION AF351812

VERSION AF351812.1 GI:15146431

KEYWORDS .

SEGMENT 1 of 13

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2809)

AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H., Ose,L., Stalenhoef,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E., Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and Patel,S.B.

TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and

sterolin-2, encoded by ABCG5 and ABCG8, respectively
JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
MEDLINE 21344600
PUBMED 11452359
REFERENCE 2 (bases 1 to 2809)
AUTHORS Lu, K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers
source 1. .2809
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="between D2S2294 and D2S2298"
/clone="1081G2; 328I4"
/cell_type="ES cell"
exon <1227. .1289
/gene="ABCG8"
/number=1

ORIGIN

Query Match 15.4%; Score 241.2; DB 9; Length 2809;
Best Local Similarity 54.8%; Pred. No. 3.3e-63;
Matches 648; Conservative 1; Mismatches 489; Indels 44; Gaps 7;

Qy	120	CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT	179
Db	1151	CTCTGTTTCCTGGAGCAGGGACACCTCGGCCTCCTGCCCTGGGCCCGTCTCTCCCAGCAT	1092
Qy	180	TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTTCGAGATAAG	239
		:	
Db	1091	TCCTTGCTGGCAAGCCACC---TACAAACGTGTGTGTTCTTGCCCACTGTCAAGATAAG	1035
Qy	240	GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTTA	299
Db	1034	GACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCAGTCTCTGCTG	975
Qy	300	TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT	359
Db	974	TCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGGCTCCCCAAC	915
Qy	360	TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTGCATTTGCTTCCTGCTA	419
Db	914	TGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTG	855
Qy	420	GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA	479
Db	854	GCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA	795
Qy	480	GGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTA	539
Db	794	GGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTG	738
Qy	540	GGTGTCTGTCATGTGTCCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGC	599

Db 737 GGCATCCTCCATGCCTCCTACAGCGTCAGGTAAGG-----CAGAGCCCTTGC 691
 Qy 600 TCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTGTCTGTCCAG 659
 Db 690 TGCTGCTGCTCCCCAGGAGTGCGGGGCCCCGGCGCTCACCCTCTGCTGCCTTTCTTCAC 631
 Qy 660 CAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTCCT 719
 Db 630 TCTTTAAGTGCCAGTCTGGGCACTTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGG 571
 Qy 720 GCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCA 779
 Db 570 AGAGGGAGAAGGGCTGTTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCCTCCCCCTGC 511
 Qy 780 CTGATTTCTGCTCTCCCCCTTCCTTGACTCGCC-----CACCACCTGTCTGTGTAGAT 832
 Db 510 CCAGGGAAGGGCCTGGGCATAAACAAAGTGGCAGCAGTGCCCTGCCAACCCAGTGTCTAC 451
 Qy 833 GGAGAAGGCTC-----GGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGC 886
 Db 450 GGCCTGCCCTCTGTGGATGGGAATGGGGGTACTGCGAATGCAAGGAGTCTTGAAACCTGG 391
 Qy 887 TGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGT 946
 Db 390 TGAAAGAATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTTCAGAGC 337
 Qy 947 CAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTT 1006
 Db 336 AACTCCAGCACAGGCCCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTT 277
 Qy 1007 GTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCT 1066
 Db 276 CTACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCCT-----GTCCGGAGACTAC 223
 Qy 1067 GTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACC 1126
 Db 222 TGGGGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGCCCCCTTCCAGGGCC 163
 Qy 1127 CTACTCTGCCTTTGCCCTTGTGGGATTTCTTTTAAAGCAACCGTGTGCGGCCTTGGTGGA 1186
 Db 162 CCAAGCTGCCTTTGCCTTCTGCGGTTTCTTTTAAAGCCACCGCGTGAGGCCCTGGTGGA 103
 Qy 1187 ACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACA 1246
 Db 102 ACATCACATCTTGCCGGCAGCAGTGGACCAGGACAGATCCTCAAAGATGTCTCCTTGTACG 43
 Qy 1247 TCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTA 1288
 Db 42 TGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGTA 1

RESULT 24

AC146464

LOCUS AC146464 202533 bp DNA linear HTG 19-AUG-2003

DEFINITION Saimiri sciureus clone CH254-84A11, WORKING DRAFT SEQUENCE.

ACCESSION AC146464

VERSION AC146464.1 GI:33636782
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
 SOURCE Saimiri sciureus (common squirrel monkey)
 ORGANISM Saimiri sciureus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;
 Saimiri.
 REFERENCE 1 (bases 1 to 202533)
 AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
 Peng, Z., Malinov, I. and Rubin, E.M.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 202533)
 AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
 Peng, Z., Malinov, I. and Rubin, E.M.
 TITLE Direct Submission
 JOURNAL Submitted (14-AUG-2003) Genome Sciences, Lawrence Berkeley National
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
 REFERENCE 3 (bases 1 to 202533)
 AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
 Peng, Z., Malinov, I. and Rubin, E.M.
 TITLE Direct Submission
 JOURNAL Submitted (19-AUG-2003) Genome Sciences, Lawrence Berkeley National
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

COMMENT

Sequence Produced by Berkeley PGA
 Web site: <http://pga.lbl.gov>
 Center Code: PGABERK
 Center Project Name: S030
 Bac Clone Name: CH254-84A11

This sequence has been compared to sequences of other species
 using Vista (<http://www-gsd.lbl.gov/VISTA>). The results can be
 viewed at:
http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=ABCG5

The order-orientation of the draft sequence was accomplished by
 using:
 Avid (<http://baboon.math.berkeley.edu/mavid>),
 Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

Summary Statistics:

Sequencing vector: Plasmid; pUC18
 Chemistry: Dye-terminator Big Dye
 Assembly program: Phrap version 0.990329.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 202533: contig of 202533 bp in length.

Qy	120	CTTTGCTCCTTAGAGCTGGGGACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCAGCAT	179
Db	32371	CTCTGTTTCTTGAGCAGGGACGCTTCAGCCTCCTGCCCTGGGTCTGGCTCTCCAGCAT	32430
Qy	180	TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTGCGAGATAAG	239
		:	
Db	32431	TCCTCTCTGGCAAGCCCA-CCTGCAAACACA-TGTGTGTTCTGCCCTCTCTCAAGATAAG	32488
Qy	240	GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTTA	299
Db	32489	GACGCGCTGGCTAAAGGTACATCAGATAACGGCCTCCTTGGCCAAGTCCTAGTCCTGCCA	32548
Qy	300	TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT	359
Db	32549	TC-----CTGAGGCTCAGGTGGAGCCAGCAGGGCAGTCTGCCACTGGCTCCCCAAC	32599
Qy	360	TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACTTGCAATTTGCTTCCTGCTA	419
Db	32600	TGCAGCCACTCCGAGGAGGGTCAGGCTACCAGAAAATCTGCTCAGCTTTGCTGCCCGTTG	32659
Qy	420	GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA	479
Db	32660	GCCATGGGTGACCTTCCATCTTTGACCCCCAGAGGGTCCATAGGACTCCAGGGAAACAGA	32719
Qy	480	GGGTCTCTGAGCTCCCTGGAGCAAGGTTCTGGTCACGGGCACAGAGGCTCGGCACAGCTTA	539
Db	32720	GGCTCCAGAGCTCCCTGGAGGGGGCTCTTGCCACTGCACCTGAGCCT---CACAGTCTG	32776
Qy	540	GGTGTCTTGCATGTGTCTTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGC	599
Db	32777	GGCGTCCTCCATGCCTCCTACAGCATCAGGTAAGGCAGAGCCCTTGCTGCTGCTCCTCCC	32836
Qy	600	TCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAG	659
Db	32837	CAGGAGCACGGGGCCCTACGTTGCGCCCTCTGCTGCCTTTTTTCACTCTTGAGCTGCCTGG	32896
Qy	660	CAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTCCT	719
Db	32897	CTGGGGACTTTGGGCTCCCTCTTCAGTGGATCGGATGGAGAGAAGAGAGCAGGGAGGGCT	32956
Qy	720	GCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCT--	777
Db	32957	GCACTGGGAAATAGGTAGCAACAGTAAATGGCTCCTCCCTCTGCCCAGGGAAGGGCCTGG	33016
Qy	778	-----CACTGATTCTGCTCTCCCTTCCTTGACTC--GCCACCACCTGTCTGTGTA	829

Db	33017	TAATAAACAAAGTTGCAGCTGTGCCCTGCCTACCCAGTGTCCACCGCTTGCCCTCTGCG	33076
Qy	830	GATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGA	889
Db	33077	GATGGAGAGAATCTGGGAATGGGG--CCTGGGAATGCAAGGAGTCTTGAATCCAGGTG-	33133
Qy	890	AGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAG	949
Db	33134	ACGAATGCAGGGACAACCACCTCCCAGACACATGGGCAGGACATTCCGAGCAGCTCCAGC	33193
Qy	950	AGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTC	1009
Db	33194	ACAGGCCCCCTTCCCTAGGAGACAGACAGCCTCAGTCGCTACCTGCCAGGTTCTACAGAGG	33253
Qy	1010	ATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTC	1069
Db	33254	ATGGAGGCTGAAACACAACACGTTAGGAGCCTGTCTGAAGATAACTGGGGT-----	33304
Qy	1070	CAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTA	1129
Db	33305	----GGCACACAGGTGGGATCAATGCTGGGGACCTGGGTGTAGCCCCCTTCCAGGGCCCCA	33360
Qy	1130	CTCTGCCTTTGCCCTTGTGGGATTTCCCTTTAAAGCAACCGTGTGGGCCTTGGTGGAACA	1189
Db	33361	TGCTGCCTTTGCCCTTCCCTGGGATTTCCCTTTAAAGCCACCGTGTGGAGCCCTGGTGGGACA	33420
Qy	1190	TCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCG	1249
Db	33421	TCACATCTTGCCGGCGACAGTGGACCAGGCAGATCCTCAAAGACGTCTCCTTATATGTGG	33480
Qy	1250	AGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGC	1309
Db	33481	AGAGCGGGCAGGTCATGTGCATCCTAGGAAGCTCAGGTAAG-----CTTGGGATGAAGGA	33535
Qy	1310	TCCTGTACTTCTAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAG	1369
Db	33536	TTCTGAA-----AAGGCTTTGGCTTGAGTTAAACTCCACCCTGAAG	33576
Qy	1370	AAGTRAGTTTAAGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATCTGGTGA	1429
Db	33577	AA-ACAGATAGATTTGTAGCAAGAAAGCCACAGGTTTGATATTAGAATGAAATCTAATGA	33635
Qy	1430	CTTGTCTGGCTGCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAAATCCCTGGAA	1489
Db	33636	--TGTCTGACTGTTAATAGAACCTGCCACCAATGTGAAATCTATAGCAAGAT-CCTTGAA	33692
Qy	1490	AGAGTGGAAAGTCCTGCCTAACACGTAAGTGCCTTCTT	1527
Db	33693	AGATTATAAAAATCCTGCCTAACACATACGTGAATTCAT	33730

RESULT 25

AX747300

LOCUS AX747300 2512 bp mRNA linear PAT 20-JUN-2003

DEFINITION Sequence 825 from Patent EP1308459.

ACCESSION AX747300

VERSION AX747300.1 GI:32131688
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
 Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
 Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
 Masuho,Y.
 TITLE Full-length cDNA sequences
 JOURNAL Patent: EP 1308459-A 825 07-MAY-2003;
 Helix Research Institute (JP) ; Research Association for
 Biotechnology (JP)
 FEATURES Location/Qualifiers
 source 1. .2512
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 ORIGIN
 Query Match 13.7%; Score 215; DB 6; Length 2512;
 Best Local Similarity 54.5%; Pred. No. 5e-55;
 Matches 576; Conservative 0; Mismatches 450; Indels 31; Gaps 6;
 Qy 237 AAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTTGGTGAAGT 296
 ||||| | |||||||||||||||||||| || ||||| | ||
 Db 1 AAGGACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAAGTCCCATTCCTG 60
 Qy 297 TTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTC 356
 | || || | | | | | | | | | | | | | | | ||||| |
 Db 61 CTGTCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGGTCCCC 120
 Qy 357 AGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCATTTGCTTCCTG 416
 | | |||| | |||| | || | ||||| ||||| | ||||| ||||
 Db 121 AACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTG 180
 Qy 417 CTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAAC 476
 | ||||||||| || | | || | | | | | | | | | | |
 Db 181 TTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAAC 240
 Qy 477 AGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGC 536
 ||||| || | ||||||||||| | | | | | | | | |||||
 Db 241 AGAGGCTCCAGAGCTCCCTGGAGGGGGTCCCTGCCACCGCCCCGAGCCT---CACAGC 297
 Qy 537 TTAGGTGTCTGCATGTGTCTTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTA 596
 | || |||| |||| | |||||||||||||||| || ||||
 Db 298 CTGGGCATCCTCCATGCCTCCTACAGCGTCAGGTAAGGCAGAGCCC-----TTGCTG 349
 Qy 597 GGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTC 656
 || | | | | | | | | | | | | | | | | | |
 Db 350 CTGCTGCTCCCCCAGGAGTGCGGGGCCCGGCGCTACCCCTCTGCTGCCTTTCTTCACTC 409
 Qy 657 CAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCT 716
 || | | | | | | | | | | | | | |
 Db 410 TTTAAGTGCCAGTCTGGGCACTTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGGAG 469

Qy 717 CCTGCGCAGGGCAGAGCCTGGACATTAACATGCCCTGCCTGAAGCCGCTTGCTGCTTC 776
 | | | | | | | | | | | | | | | | | | | | | |
 Db 470 AGGGAGAAGGGCTGTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCCTCCCCCTGCCCA 529
 Qy 777 TCACTGATTCTGCTCTCCCCTTCCTTGACTC-GCCCACCACCTGTCTGTGTAGATGGA 835
 | | | | | | | | | | | | | | | | | | | | | |
 Db 530 GGGAAAGGGCCTGGGCATAAACAAAGTGGCAGCAGTGCCCTGCCAACCCAGTGTCTACGGC 589
 Qy 836 GAAGGCTC-----GGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGA 889
 | | | | | | | | | | | | | | | | | | | | | |
 Db 590 CTGCCCTCTGTGGATGGGAATGGGGGTACTGCGAATGCAAGGAGTCTTGAAACCTGGTGA 649
 Qy 890 AGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCCAGTCAG 949
 | | | | | | | | | | | | | | | | | | | | | |
 Db 650 AAGAATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTCAGAGCAAC 703
 Qy 950 AGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGCCAACTTGTC 1009
 | | | | | | | | | | | | | | | | | | | | | |
 Db 704 TCCAGCACAGGCCCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTCT- 762
 Qy 1010 ATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTC 1069
 | | | | | | | | | | | | | | | | | | | | | |
 Db 763 -----ACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCCTGTCCGGAGACTACTG 816
 Qy 1070 CAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTA 1129
 | | | | | | | | | | | | | | | | | | | | | |
 Db 817 GGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCCCTCCAGGGCCCCA 876
 Qy 1130 CTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGCCTTGGTGGGAACA 1189
 | | | | | | | | | | | | | | | | | | | | | |
 Db 877 AGCTGCCTTTGCCTTCCTGGGGTTTCTTTAAAGCCACCGCGTGAGGCCCTGGTGGGACA 936
 Qy 1190 TCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCG 1249
 | | | | | | | | | | | | | | | | | | | | | |
 Db 937 TCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGG 996
 Qy 1250 AGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGG 1286
 | | | | | | | | | | | | | | | | | | | | | |
 Db 997 AGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGG 1033

RESULT 26

AK091997

LOCUS AK091997 2512 bp mRNA linear PRI 15-JUL-2002

DEFINITION Homo sapiens cDNA FLJ34678 fis, clone LIVER2003065.

ACCESSION AK091997

VERSION AK091997.1 GI:21750490

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
 Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
 Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,


```

Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A.,
Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K.,
Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y.,
Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE      NEDO human cDNA sequencing project
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 2512)
AUTHORS    Isogai,T. and Yamamoto,J.
TITLE      Direct Submission
JOURNAL    Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT    NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
  source          Location/Qualifiers
                  1. .2512
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="LIVER2003065"
                  /tissue_type="liver"
                  /clone_lib="LIVER2"
                  /note="cloning vector: pME18SFL3"
ORIGIN
Query Match          13.7%;  Score 215;  DB 9;  Length 2512;
Best Local Similarity 54.5%;  Pred. No. 5e-55;
Matches 576;  Conservative 0;  Mismatches 450;  Indels 31;  Gaps 6;

Qy      237  AAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGT 296
      ||||| | ||||| ||||| ||||| || ||||| | |||||
Db      1   AAGGACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCATTTCCTG 60

Qy      297  TTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTC 356
      | ||| | | | | | | | | | | | | | | | | | ||||| |
Db      61  CTGTCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCC 120

Qy      357  AGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAGTTGCATTTGCTTCCTG 416
      | | |||| | |||| | || | ||||| ||||| | | ||||| ||||
Db      121  AACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTG 180

Qy      417  CTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGGAGCCAGAGGGCCTCACATCAAC 476
      | ||||| ||||| || | | | || | | | || | || | |||
Db      181  TTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAAC 240

Qy      477  AGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGC 536
      ||||| || | ||||| ||||| | | | || | ||||| |||||
Db      241  AGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGC 297

```


ACCESSION AX320881
 VERSION AX320881.1 GI:17902431
 KEYWORDS .
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Tian, H., Schultz, J. and Shan, B.
 TITLE Sitosterolemia susceptibility gene (ssg): compositions and methods
 of use
 JOURNAL Patent: WO 0179272-A 2 25-OCT-2001;
 Tularik Inc. (US)
 FEATURES Location/Qualifiers
 source 1. .2258
 /organism="Mus musculus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:10090"
 /note="mouse sitosterolemia susceptibility gene (SSG)"
 CDS 47. .2005
 /note="unnamed protein product; mouse sitosterolemia
 susceptibility gene (SSG) protein"
 /codon_start=1
 /protein_id="CAD19408.1"
 /db_xref="GI:17902432"
 /db_xref="REMTREMBL:CAD19408"
 /translation="MGELPFLSPEGARGPHINRGSLSSEQGSVTGTEARHSLGVLHV
 SYSVSNRVGPWWNIKSCQQKWDRLKDVSLYIESGQIMCILGSSSGSKTLLDAISG
 RLRTGTLEGEVFNVCGLRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS
 SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGSISSGERRRVSIAAQLLQDPKVMMLD
 EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG
 TPEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTSREREIETIKRVQMLECAFKESD
 IYHKILENIERARYLKTLPMPVFKTKDPPGMFGKLGVLRLRVTRNLMRNKQAVIMRLV
 QNLIMGLFLIFYLLRVQNNLTKGAVQDRVGLLYQLVGATPYTGMNLAVNLFPMRLAVS
 DQESQDGLYHKWQMLLAYVLHVLPPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALL
 APHLIGEFLTLVLLGIVQNPINIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
 TFQKYCCEILVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN
 FLILYGFIPALVILGIVIFKVRDYILIS"

ORIGIN

Query Match 12.2%; Score 191.4; DB 6; Length 2258;
 Best Local Similarity 97.0%; Pred. No. 1.2e-47;
 Matches 195; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy	378	GGACAGGCCACTAGAAAATTCACCTTGCACTTGCTTCCTGCTAGCCATGGGTGAGCTGCCC	437
Db	2	GGACAGGCCACTAGAAAATTCACCTTGCACTTGCTTCCTGCTAGCCATGGGTGAGCTGCCC	61
Qy	438	TTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTG	497
Db	62	TTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTG	121
Qy	498	GAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCC	557
Db	122	GAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCC	181
Qy	558	TACAGCGTCAGGTAAGGGGAC	578

||||||| | | |
Db 182 TACAGCGTCAGCAACCGTGTC 202

RESULT 28

AF404107/c

LOCUS AF404107 581 bp DNA linear PRI 14-AUG-2001

DEFINITION Homo sapiens sterolin 1 (ABCG5) and sterolin 2 (ABCG8) genes,
partial cds.

ACCESSION AF404107

VERSION AF404107.1 GI:15150318

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 581)

AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H.,
Ose,L., Stalenhoef,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E.,
Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and
Patel,S.B.

TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterolin-1 and
sterolin-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)

MEDLINE 21344600

PUBMED 11452359

REFERENCE 2 (bases 1 to 581)

AUTHORS Lu,K., Lee,M. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (11-JUL-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403

FEATURES Location/Qualifiers

source

1. .581
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

gene

complement(<1. .>143)
/gene="ABCG5"

mRNA

complement(<1. .>143)
/gene="ABCG5"

CDS

/product="sterolin 1"
complement(<1. .143)
/gene="ABCG5"
/codon_start=1
/product="sterolin 1"
/protein_id="AAK85388.1"
/db_xref="GI:15150319"
/translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
YSV"

exon

complement(<1. .>143)
/gene="ABCG5"
/number=1

misc_feature

144. .518
/note="contains 5'UTR and promoter regions for ABCG5 and
ABCG8"

```

gene          <519. .>581
               /gene="ABCG8"
mRNA          <519. .>581
               /gene="ABCG8"
               /product="sterolin 2"
CDS           519. .>581
               /gene="ABCG8"
               /codon_start=1
               /product="sterolin 2"
               /protein_id="AAK85389.1"
               /db_xref="GI:15150320"
               /translation="MAGKAAEERGLPKGATPQDTS"
exon          <519. .>581
               /gene="ABCG8"
               /number=1

```

ORIGIN

```

Query Match          11.5%;  Score 179.8;  DB 9;  Length 581;
Best Local Similarity 65.5%;  Pred. No. 4.6e-44;
Matches 294;  Conservative 1;  Mismatches 148;  Indels 6;  Gaps 2;

```

```

Qy      120 CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
          || || | ||| ||| |||| ||| | | | | | | | | | | | | | | | | |
Db      443 CTCTGTTTCTTGGAGCAGGGACACCTCGGCCTCCTGCCCTGGGCCCGTCTCTCCCAGCAT 384

Qy      180 TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAG 239
          ||||: ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db      383 TCCTTGCTGGCAAGCCCACC---TACAAACGTGTGTGTTCTTGCCCACTGTCAAGATAAG 327

Qy      240 GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACGTGTTA 299
          ||| | ||||| ||||| ||||| || ||||| || ||||| || ||||| || |||||
Db      326 GACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAAGTCCCAGTCTTGCTG 267

Qy      300 TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGT 359
          || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      266 TCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAAC 207

Qy      360 TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTGCATTTGCTTCCTGCTA 419
          | |||| | |||| | || | ||||| ||||| | | | | | | | | | | | |
Db      206 TGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTG 147

Qy      420 GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA 479
          ||||| ||||| || | | | | | | | | | | | | | | | | | | | | | |
Db      146 GCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA 87

Qy      480 GGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTA 539
          || || | ||||| ||||| | | | | | | | | | | | | | | | | | | |
Db      86 GGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTG 30

Qy      540 GGTGTCCTGCATGTGTCCTACAGCGTCAG 568
          || |||| |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      29 GGCATCCTCCATGCCTCCTACAGCGTCAG 1

```

RESULT 29

AC084712

LOCUS AC084712 68166 bp DNA linear HTG 08-NOV-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-328I4 map 2, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC084712
 VERSION AC084712.1 GI:11120851
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 68166)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 2, clone RP11-328I4
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 68166)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (08-NOV-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11395
 Center clone name: 328_I_4

* NOTE: This record contains 83 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone

* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1	734: contig of 734 bp in length
* 735	834: gap of 100 bp
* 835	1564: contig of 730 bp in length
* 1565	1664: gap of 100 bp
* 1665	2396: contig of 732 bp in length
* 2397	2496: gap of 100 bp
* 2497	3211: contig of 715 bp in length
* 3212	3311: gap of 100 bp
* 3312	4013: contig of 702 bp in length
* 4014	4113: gap of 100 bp
* 4114	4817: contig of 704 bp in length
* 4818	4917: gap of 100 bp
* 4918	5617: contig of 700 bp in length
* 5618	5717: gap of 100 bp
* 5718	6384: contig of 667 bp in length
* 6385	6484: gap of 100 bp
* 6485	7189: contig of 705 bp in length
* 7190	7289: gap of 100 bp
* 7290	8030: contig of 741 bp in length
* 8031	8130: gap of 100 bp
* 8131	8876: contig of 746 bp in length
* 8877	8976: gap of 100 bp
* 8977	9688: contig of 712 bp in length
* 9689	9788: gap of 100 bp
* 9789	10495: contig of 707 bp in length
* 10496	10595: gap of 100 bp
* 10596	11330: contig of 735 bp in length
* 11331	11430: gap of 100 bp
* 11431	12167: contig of 737 bp in length
* 12168	12267: gap of 100 bp
* 12268	13007: contig of 740 bp in length
* 13008	13107: gap of 100 bp
* 13108	13844: contig of 737 bp in length
* 13845	13944: gap of 100 bp
* 13945	14628: contig of 684 bp in length
* 14629	14728: gap of 100 bp
* 14729	15453: contig of 725 bp in length
* 15454	15553: gap of 100 bp
* 15554	16285: contig of 732 bp in length
* 16286	16385: gap of 100 bp
* 16386	17111: contig of 726 bp in length
* 17112	17211: gap of 100 bp
* 17212	17905: contig of 694 bp in length
* 17906	18005: gap of 100 bp
* 18006	18711: contig of 706 bp in length
* 18712	18811: gap of 100 bp
* 18812	19500: contig of 689 bp in length
* 19501	19600: gap of 100 bp
* 19601	20319: contig of 719 bp in length
* 20320	20419: gap of 100 bp
* 20420	21149: contig of 730 bp in length
* 21150	21249: gap of 100 bp
* 21250	21936: contig of 687 bp in length
* 21937	22036: gap of 100 bp

*	22037	22775: contig of 739 bp in length
*	22776	22875: gap of 100 bp
*	22876	23594: contig of 719 bp in length
*	23595	23694: gap of 100 bp
*	23695	24434: contig of 740 bp in length
*	24435	24534: gap of 100 bp
*	24535	25259: contig of 725 bp in length
*	25260	25359: gap of 100 bp
*	25360	26076: contig of 717 bp in length
*	26077	26176: gap of 100 bp
*	26177	26885: contig of 709 bp in length
*	26886	26985: gap of 100 bp
*	26986	27521: contig of 536 bp in length
*	27522	27621: gap of 100 bp
*	27622	28356: contig of 735 bp in length
*	28357	28456: gap of 100 bp
*	28457	29193: contig of 737 bp in length
*	29194	29293: gap of 100 bp
*	29294	30023: contig of 730 bp in length
*	30024	30123: gap of 100 bp
*	30124	30866: contig of 743 bp in length
*	30867	30966: gap of 100 bp
*	30967	31708: contig of 742 bp in length
*	31709	31808: gap of 100 bp
*	31809	32546: contig of 738 bp in length
*	32547	32646: gap of 100 bp
*	32647	33374: contig of 728 bp in length
*	33375	33474: gap of 100 bp
*	33475	34213: contig of 739 bp in length
*	34214	34313: gap of 100 bp
*	34314	35026: contig of 713 bp in length
*	35027	35126: gap of 100 bp
*	35127	35837: contig of 711 bp in length
*	35838	35937: gap of 100 bp
*	35938	36632: contig of 695 bp in length
*	36633	36732: gap of 100 bp
*	36733	37434: contig of 702 bp in length
*	37435	37534: gap of 100 bp
*	37535	38250: contig of 716 bp in length
*	38251	38350: gap of 100 bp
*	38351	39076: contig of 726 bp in length
*	39077	39176: gap of 100 bp
*	39177	39911: contig of 735 bp in length
*	39912	40011: gap of 100 bp
*	40012	40753: contig of 742 bp in length
*	40754	40853: gap of 100 bp
*	40854	41564: contig of 711 bp in length
*	41565	41664: gap of 100 bp
*	41665	42413: contig of 749 bp in length
*	42414	42513: gap of 100 bp
*	42514	43254: contig of 741 bp in length
*	43255	43354: gap of 100 bp
*	43355	44101: contig of 747 bp in length
*	44102	44201: gap of 100 bp
*	44202	44944: contig of 743 bp in length
*	44945	45044: gap of 100 bp
*	45045	45757: contig of 713 bp in length


```

* 45758 45857: gap of 100 bp
* 45858 46590: contig of 733 bp in length
* 46591 46690: gap of 100 bp
* 46691 47437: contig of 747 bp in length
* 47438 47537: gap of 100 bp
* 47538 48287: contig of 750 bp in length
* 48288 48387: gap of 100 bp
* 48388 49095: contig of 708 bp in length
* 49096 49195: gap of 100 bp
* 49196 49939: contig of 744 bp in length
* 49940 50039: gap of 100 bp
* 50040 50778: contig of 739 bp in length
* 50779 50878: gap of 100 bp
* 50879 51619: contig of 741 bp in length
* 51620 51719: gap of 100 bp
* 51720 52438: contig of 719 bp in length
* 52439 52538: gap of 100 bp
* 52539 53235: contig of 697 bp in length
* 53236 53335: gap of 100 bp
* 53336 54028: contig of 693 bp in length
* 54029 54128: gap of 100 bp
* 54129 54853: contig of 725 bp in length
* 54854 54953: gap of 100 bp
* 54954 55679: contig of 726 bp in length
* 55680 55779: gap of 100 bp
* 55780 56519: contig of 740 bp in length
* 56520 56619: gap of 100 bp
* 56620 57358: contig of 739 bp in length

```

```

Query Match          11.1%; Score 174; DB 2; Length 68166;
Best Local Similarity 64.2%; Pred. No. 3.6e-42;
Matches 292; Conservative 1; Mismatches 157; Indels 5; Gaps 2;

```

```

Qy      120 CTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 179
      || || | | | |||| ||| ||||| ||| | | | | | | | | | | | | | | | | |
Db      34526 CTCTGTTTCCTGGAGCAGGGACACCTCGGCCTCCTGCCCTGGGCCCCTCTCTCCCAGCAT 34585

Qy      180 TCCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAG 239
      ||||: ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      34586 TCCTTGCTGGCAAGCCACCTACAAACGT----GTGTGTTCTGCCCACTGTCAAGATAAG 34641

Qy      240 GACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTTA 299
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      34642 GACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCACTCCTGCTG 34701

Qy      300 TCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGT 359
      || || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      34702 TCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAAC 34761

Qy      360 TAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAAATTCACCTTGCAATTTGCTTCCTGCTA 419
      | |||| | |||| | || | ||||| | ||||| | | | | | | | | | | | | | |
Db      34762 TGAAGCCACTCTGGGGAGGGTCCGGCCACCAAAAAATTTGCCCAGCTTTGCTGCCTGTTG 34821

Qy      420 GCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGA 479
      ||||| ||||| || | | | || | || | || | || | || | || | || | || | ||
Db      34822 GCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA 34881

```

Qy 480 GGGTCTCTGAGCTCCCTGGAGCAAGGTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTA 539
 | | | | | | | | | | | | | | | | | | | | | | |
Db 34882 AGCTCCCAGAGCTCCCTGG-GAGGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTG 34940

Qy 540 GGTGTCCTGCATGTGTCTTACAGCGTCAGGTAAGG 574
 || |||| |||| ||||| | | | | | | | |
Db 34941 GGCATCCTCCATGCCTCCTACAGCGTNAAGTAAGG 34975

RESULT 30

AF312714

LOCUS AF312714 2470 bp mRNA linear ROD 26-AUG-2002

DEFINITION *Rattus norvegicus* sterolin (Abcg5) mRNA, complete cds.

ACCESSION AF312714

VERSION AF312714.3 GI:22477143

KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 2470)

AUTHORS Lee,M.H., Lu,K., Hazard,S., Yu,H., Shulenin,S., Hidaka,H.,
Kojima,H., Allikmets,R., Sakuma,N., Pegoraro,R., Srivastava,A.K.,
Salen,G., Dean,M. and Patel,S.B.

TITLE Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption

JOURNAL Nat. Genet. 27 (1), 79-83 (2001)

MEDLINE 20578753

PUBMED 11138003

REFERENCE 2 (bases 1 to 2470)

AUTHORS Lu, K., Lee, M.-H. and Patel, S.B.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

REFERENCE 3 (bases 1 to 2470)

AUTHORS Lu, K., Lee, M.-H. and Patel, S.B.

TITLE Direct Submission

JOURNAL Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

REMARK Sequence update by submitter

REFERENCE 4 (bases 1 to 2470)

AUTHORS Lu, K., Lee, M. and Patel, S.B.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-2002) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St. STB 541, Charleston, SC 29403, USA

REMARK Sequence update by submitter

COMMENT On Aug 26, 2002 this sequence version replaced gi:14091945.

FEATURES	Location/Qualifiers
----------	---------------------

source

1. .2470

```
/organism="Rattus norvegicus"
```

```
/mol type="mRNA"
```

```
/strain="Sprague-Dawley"
```

```
/db xref="taxon:10116"
```

```

/tissue_type="liver"
gene      1. .2470
          /gene="Abcg5"
CDS       65. .2023
          /gene="Abcg5"
          /note="ABCG5"
          /codon_start=1
          /product="sterolin"
          /protein_id="AAG53098.3"
          /db_xref="GI:22477144"
          /translation="MSELPFLSPEGARGPHNNRGSQSSLEEGSVTGSEARHSLGVLNV
SFSVSNRVGPWWNIKSCQQKWDRLKLDVSLYIESGQTMCILGSSSGSKTTLDDAISG
RLRRTGTLEGEVFNVCGLRRDQFQDCVSYLLQSDVFLSSLTVRETLRYTAMLALRSS
SADFYDKKVEAVLTELSLSHVADQMIGNYNFGGISSGERRRVSIQAQLLQDPKVMMLD
EPTTGLDCMTANHIVLLLVELARRNRIVITIHQPRSELFHHFDKIAILTYGELVFCG
TPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLESAFRQSD
ICKILENIERTRHLKTLPMVPFKTKNPPGMFCKLGVLLRRVTRNLMRNKQVVIMRLV
QNLIMGLFLIFYLLRVQNNMLKGAQVDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVS
DQESQDGLYQKWQMLLAYVLHALPFSIVATVIFSSVCYWTGLGYPEVARFGYFSAALL
APHLIGEFLTLVLLGMVQNPINIVNSIVALLSISGLLIGSGFIRNIEEMPIPLKILGYF
TFQKYCCILVNEFYGLNFTCGSNTSVPNNPMCSMTQGIQFIEKTCPGATSRFTTN
FLILYSFIPTLVILGMVVFKVRDYLIIR"

```

ORIGIN

```

Query Match      11.1%;  Score 173.6;  DB 10;  Length 2470;
Best Local Similarity  86.8%;  Pred. No. 4.2e-42;
Matches 191;  Conservative  0;  Mismatches  29;  Indels  0;  Gaps  0;

```

```

Qy      359  TTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCATTGCAATTTGCTTCCTGCT 418
          ||||| ||| ||| |||| ||||| ||| ||| ||||| ||||| |||
Db      1    TTAAAGTTGCTCTGAAGCCAGACAGGACACCAGAGGATTCACTCACATTTGCTTCCCGCT 60

Qy      419  AGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAG 478
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61  GGCCATGAGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACAACAACAG 120

Qy      479  AGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTT 538
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121  AGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTCAGAGGCTCGGCACAGCTT 180

Qy      539  AGGTGTCCTGCATGTGTCTTACAGCGTCAGGTAAGGGGAC 578
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      181  AGGTGTCCTGAATGTGTCTTACAGCGTCAGCAACCGTGTC 220

```

RESULT 31

AY196216/c

LOCUS AY196216 2284 bp mRNA linear ROD 01-JUN-2003

DEFINITION Mus musculus strain PERA/Ei ATP-binding cassette sub-family G member 8 (Abcg8) mRNA, complete cds.

ACCESSION AY196216

VERSION AY196216.1 GI:31322261

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Qy 121 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAG 164
 ||||||||||||||||||||||||||||||||||||||||
 Db 44 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAG 1

RESULT 32
 AY196215/c

LOCUS AY196215 2285 bp mRNA linear ROD 01-JUN-2003
 DEFINITION Mus musculus strain I/LnJ ATP-binding cassette sub-family G member
 8 (Abcg8) mRNA, complete cds.

ACCESSION AY196215
 VERSION AY196215.1 GI:31322259
 KEYWORDS .

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2285)
 AUTHORS Wittenburg,H., Lyons,M.A., Li,R., Churchill,G.A., Carey,M.C. and
 Paigen,B.
 TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone
 Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred
 Mice

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2285)
 AUTHORS Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.
 TITLE Direct Submission
 JOURNAL Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street,
 Bar Harbor, ME 04609, USA

FEATURES Location/Qualifiers
 source 1. .2285
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="I/LnJ"
 /db_xref="taxon:10090"
 /chromosome="17"
 /map="55 cM"
 /sex="male"
 /tissue_type="liver"
 gene 1. .2285
 /gene="Abcg8"
 CDS 102. .2120
 /gene="Abcg8"
 /note="ATP-dependent canalicular cholesterol transporter;
 white subfamily"
 /codon_start=1
 /product="ATP-binding cassette sub-family G member 8"
 /protein_id="AAO45095.1"
 /db_xref="GI:31322260"
 /translation="MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQS
 NTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLA
 IIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPN
 LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
 RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDI
 FRLFDLVLLMTSGTP IYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKER
 EVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTLTQDTCGTAAELP

GMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAAL
 LFMIGALIPFNVILDVVSCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVI
 IYAMPIYWLTNLRPVPELFLHLLLVWLVVFCRTMALAASAMLPTFHMSSFFCNALY
 NSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSIL
 GDTMISAMDNLNHPYAIYLVIGISYGFLEFLYYLSLKLKQKSIQDW"

ORIGIN

Query Match 10.4%; Score 164; DB 10; Length 2285;
 Best Local Similarity 100.0%; Pred. No. 4.2e-39;
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 164 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 105

Qy 61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 104 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 45

Qy 121 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAG 164
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 44 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAG 1

RESULT 33

AF324495/c

LOCUS AF324495 3674 bp mRNA linear ROD 07-AUG-2001
 DEFINITION Mus musculus sterolin-2 (Abcg8) mRNA, complete cds.
 ACCESSION AF324495
 VERSION AF324495.1 GI:15088541
 KEYWORDS .
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 3674)
 AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H.,
 Ose,L., Stalenhoef,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E.,
 Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and
 Patel,S.B.
 TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterolin-1 and
 sterolin-2, encoded by ABCG5 and ABCG8, respectively
 JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
 MEDLINE 21344600
 PUBMED 11452359
 REFERENCE 2 (bases 1 to 3674)
 AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 Street, STB541, Charleston, SC 29403, USA
 FEATURES Location/Qualifiers
 source 1..3674
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"

```

/db_xref="taxon:10090"
/tissue_type="liver"
gene      1. .3674
          /gene="Abcg8"
CDS       102. .2123
          /gene="Abcg8"
          /note="ABCG8"
          /codon_start=1
          /product="sterolin-2"
          /protein_id="AAK84079.1"
          /db_xref="GI:15088542"
          /translation="MAEKTKEETQLWNGTVLQDASQGLQDSLFSSES DNSLYFTYSGQ
SNTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQML
AIIGSSGCGRASLLDVITGRGHGGMKSGQIWINGQPSTPQLVRKCVAHVRQHDQLLP
NLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGE
RRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSD
IFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKE
REVATVEKAQSLAALFLEKVQGFDDFLWKAELNTSTHTVSLTLTQDTCGTAVEL
PGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACIMSLIIGFLYGHGAKQLSFMDTAA
LLFMIGALIPFNVILDVSKCHSERSMLYEELEDGLYTAGPYFFAKILGELPEHCAYV
IIYAMPIYWLTLNLRPVPELFLHFLVWL VVFCRTMALAASAMLPTFMSSFFCNAL
YNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSI
LGDTMISAMDLNSHPLYAIYLIVIGISYGFLLFYLLSLKLIKQKSIQDW"

```

ORIGIN

```

Query Match      10.4%;  Score 164;  DB 10;  Length 3674;
Best Local Similarity 100.0%;  Pred. No. 4.3e-39;
Matches 164;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
        |||
Db      164 CGAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 105

Qy      61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
        |||
Db      104 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 45

Qy      121 TTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAG 164
        |||
Db      44 TTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAG 1

```

RESULT 34

BD223287

LOCUS BD223287 226 bp DNA linear PAT 17-JUL-2003

DEFINITION Toxicological response markers.

ACCESSION BD223287

VERSION BD223287.1 GI:33033057

KEYWORDS JP 2002523112-A/24.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 226)

AUTHORS Cunningham,M.J., Zweiger,G.B., Panzer,S.R. and Seilhamer,J.J.

TITLE Toxicological response markers

Qy 483 TCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 542
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 TCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120

Qy 543 GTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGAC 578
 |||||||||||||||||||| | | | |
 Db 121 GTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTC 156

RESULT 37

AX685729

LOCUS AX685729 1959 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 1 from Patent WO02081691.

ACCESSION AX685729

VERSION AX685729.1 GI:29371738

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 1 17-OCT-2002;
 Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
 (US)

FEATURES

source

Location/Qualifiers

1. .1959

/organism="Mus musculus"

/mol_type="unassigned DNA"

/db_xref="taxon:10090"

CDS

1. .1959

/note="unnamed protein product; ABCG5 (mABCG5)"

/codon_start=1

/protein_id="CAD86570.1"

/db_xref="GI:29371739"

/db_xref="REMTREMBL:CAD86570"

/translation="MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHV
 SYSVSNRVGPWWNIKSCQQKWDRLKDVSLYIESGQIMCILGSSGSGKTTLLDAISG
 RLRRRTGTLEGEVFVNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS
 SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLD
 EPTTGLDCMTANQIVLLLAELARRDRIVITIHQPRSELFQHFQDKIAILTYGELVFCG
 TPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETKYRVQMLECAFKESD
 IYHKILENIERARYLKTLPMPVFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLV
 QNLIMGLFLIFYLLRVQNNLTKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVS
 DQESQDGLYHKWQMLLAYVLHVLVPSVIATVIFSSVCYWTGLGLYPEVAREFGYFSAALL
 APHLIGEFLTLVLLGIVQNPINIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
 TFQKYCCEILVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN
 FLILYGFIPALVILGIVIFKVRDYILSR"

ORIGIN

Query Match 9.3%; Score 146.4; DB 6; Length 1959;
 Best Local Similarity 96.2%; Pred. No. 1.3e-33;
 Matches 150; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 423 ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 482

```

|||||
Db      1 ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGCCTCACATCAACAGAGGG 60
Qy      483 TCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 542
|||||
Db      61 TCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120
Qy      543 GTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGAC 578
|||||
Db      121 GTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTC 156

```

RESULT 38

AC084712/c

LOCUS AC084712 68166 bp DNA linear HTG 08-NOV-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-328I4 map 2, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC084712

VERSION AC084712.1 GI:11120851

KEYWORDS HTG; HTGS_PHASE0.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 68166)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 2, clone RP11-328I4

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 68166)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (08-NOV-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11395

Center clone name: 328_I_4

-----.

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1	734: contig of 734 bp in length
* 735	834: gap of 100 bp
* 835	1564: contig of 730 bp in length
* 1565	1664: gap of 100 bp
* 1665	2396: contig of 732 bp in length
* 2397	2496: gap of 100 bp
* 2497	3211: contig of 715 bp in length
* 3212	3311: gap of 100 bp
* 3312	4013: contig of 702 bp in length
* 4014	4113: gap of 100 bp
* 4114	4817: contig of 704 bp in length
* 4818	4917: gap of 100 bp
* 4918	5617: contig of 700 bp in length
* 5618	5717: gap of 100 bp
* 5718	6384: contig of 667 bp in length
* 6385	6484: gap of 100 bp
* 6485	7189: contig of 705 bp in length
* 7190	7289: gap of 100 bp
* 7290	8030: contig of 741 bp in length
* 8031	8130: gap of 100 bp
* 8131	8876: contig of 746 bp in length
* 8877	8976: gap of 100 bp
* 8977	9688: contig of 712 bp in length
* 9689	9788: gap of 100 bp
* 9789	10495: contig of 707 bp in length
* 10496	10595: gap of 100 bp
* 10596	11330: contig of 735 bp in length
* 11331	11430: gap of 100 bp
* 11431	12167: contig of 737 bp in length
* 12168	12267: gap of 100 bp
* 12268	13007: contig of 740 bp in length
* 13008	13107: gap of 100 bp
* 13108	13844: contig of 737 bp in length
* 13845	13944: gap of 100 bp
* 13945	14628: contig of 684 bp in length
* 14629	14728: gap of 100 bp
* 14729	15453: contig of 725 bp in length
* 15454	15553: gap of 100 bp
* 15554	16285: contig of 732 bp in length

*	16286	16385: gap of 100 bp
*	16386	17111: contig of 726 bp in length
*	17112	17211: gap of 100 bp
*	17212	17905: contig of 694 bp in length
*	17906	18005: gap of 100 bp
*	18006	18711: contig of 706 bp in length
*	18712	18811: gap of 100 bp
*	18812	19500: contig of 689 bp in length
*	19501	19600: gap of 100 bp
*	19601	20319: contig of 719 bp in length
*	20320	20419: gap of 100 bp
*	20420	21149: contig of 730 bp in length
*	21150	21249: gap of 100 bp
*	21250	21936: contig of 687 bp in length
*	21937	22036: gap of 100 bp
*	22037	22775: contig of 739 bp in length
*	22776	22875: gap of 100 bp
*	22876	23594: contig of 719 bp in length
*	23595	23694: gap of 100 bp
*	23695	24434: contig of 740 bp in length
*	24435	24534: gap of 100 bp
*	24535	25259: contig of 725 bp in length
*	25260	25359: gap of 100 bp
*	25360	26076: contig of 717 bp in length
*	26077	26176: gap of 100 bp
*	26177	26885: contig of 709 bp in length
*	26886	26985: gap of 100 bp
*	26986	27521: contig of 536 bp in length
*	27522	27621: gap of 100 bp
*	27622	28356: contig of 735 bp in length
*	28357	28456: gap of 100 bp
*	28457	29193: contig of 737 bp in length
*	29194	29293: gap of 100 bp
*	29294	30023: contig of 730 bp in length
*	30024	30123: gap of 100 bp
*	30124	30866: contig of 743 bp in length
*	30867	30966: gap of 100 bp
*	30967	31708: contig of 742 bp in length
*	31709	31808: gap of 100 bp
*	31809	32546: contig of 738 bp in length
*	32547	32646: gap of 100 bp
*	32647	33374: contig of 728 bp in length
*	33375	33474: gap of 100 bp
*	33475	34213: contig of 739 bp in length
*	34214	34313: gap of 100 bp
*	34314	35026: contig of 713 bp in length
*	35027	35126: gap of 100 bp
*	35127	35837: contig of 711 bp in length
*	35838	35937: gap of 100 bp
*	35938	36632: contig of 695 bp in length
*	36633	36732: gap of 100 bp
*	36733	37434: contig of 702 bp in length
*	37435	37534: gap of 100 bp
*	37535	38250: contig of 716 bp in length
*	38251	38350: gap of 100 bp
*	38351	39076: contig of 726 bp in length
*	39077	39176: gap of 100 bp

* 39177 39911: contig of 735 bp in length
 * 39912 40011: gap of 100 bp
 * 40012 40753: contig of 742 bp in length
 * 40754 40853: gap of 100 bp
 * 40854 41564: contig of 711 bp in length
 * 41565 41664: gap of 100 bp
 * 41665 42413: contig of 749 bp in length
 * 42414 42513: gap of 100 bp
 * 42514 43254: contig of 741 bp in length
 * 43255 43354: gap of 100 bp
 * 43355 44101: contig of 747 bp in length
 * 44102 44201: gap of 100 bp
 * 44202 44944: contig of 743 bp in length
 * 44945 45044: gap of 100 bp
 * 45045 45757: contig of 713 bp in length
 * 45758 45857: gap of 100 bp
 * 45858 46590: contig of 733 bp in length
 * 46591 46690: gap of 100 bp
 * 46691 47437: contig of 747 bp in length
 * 47438 47537: gap of 100 bp
 * 47538 48287: contig of 750 bp in length
 * 48288 48387: gap of 100 bp
 * 48388 49095: contig of 708 bp in length
 * 49096 49195: gap of 100 bp
 * 49196 49939: contig of 744 bp in length
 * 49940 50039: gap of 100 bp
 * 50040 50778: contig of 739 bp in length
 * 50779 50878: gap of 100 bp
 * 50879 51619: contig of 741 bp in length
 * 51620 51719: gap of 100 bp
 * 51720 52438: contig of 719 bp in length
 * 52439 52538: gap of 100 bp
 * 52539 53235: contig of 697 bp in length
 * 53236 53335: gap of 100 bp
 * 53336 54028: contig of 693 bp in length
 * 54029 54128: gap of 100 bp
 * 54129 54853: contig of 725 bp in length
 * 54854 54953: gap of 100 bp
 * 54954 55679: contig of 726 bp in length
 * 55680 55779: gap of 100 bp
 * 55780 56519: contig of 740 bp in length
 * 56520 56619: gap of 100 bp
 * 56620 57358: contig of 739 bp in length

Query Match 9.3%; Score 145.4; DB 2; Length 68166;
 Best Local Similarity 72.6%; Pred. No. 3.1e-33;
 Matches 207; Conservative 0; Mismatches 66; Indels 12; Gaps 1;

Qy 1019 CCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAA 1078
 ||| | | |||| | | | | | || | ||||| | | |
 Db 10159 CCAGGTTCTACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCCTGTCCGGAGACTA 10100
 Qy 1079 CTGGGATGGGG-----TAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACC 1126
 |||| | || | | |||| | || | || | ||||| ||
 Db 10099 CTGGGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCTTCCAGGGCC 10040
 Qy 1127 CTA CTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGCCCTTGGTGGA 1186


```

DEFINITION      Sequence 42 from Patent WO0227016.
ACCESSION       AX456520
VERSION         AX456520.1   GI:21715410
KEYWORDS        .
SOURCE          synthetic construct
ORGANISM        synthetic construct
                artificial sequences.
REFERENCE       1
AUTHORS         Patel,S.B. and Dean,M.
TITLE           Gene involved in dietary sterol absorption and excretion and uses
                therefor
JOURNAL         Patent: WO 0227016-A 42 04-APR-2002;
                THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
                Shailendra B. (US) ; Dean, Michael (US)
FEATURES
  source        Location/Qualifiers
                1. .2516
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="Primer"

```

Query Match 6.8%; Score 107; DB 6; Length 2516;
Best Local Similarity 63.6%; Pred. No. 2.6e-21;
Matches 180; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

Qy	294	CTGTTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTC	353
Db	12	CTGCTGTCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTC	71
Qy	354	CTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTC	413
Db	72	CCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGC	131
Qy	414	CTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATC	473
Db	132	CTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTA	191
Qy	474	AACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCTGGTCACGGGCACAGAGGCTCGGCAC	533
Db	192	AACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CAC	248
Qy	534	AGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGG	576
Db	249	AGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG	291

RESULT 41

AF312715

```
LOCUS       AF312715                2740 bp    mRNA    linear    PRI 14-JUN-2001
DEFINITION  Homo sapiens sterolin (ABCG5) mRNA, complete cds.
ACCESSION   AF312715
VERSION     AF312715.2  GI:14423628
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```


Qy 294 CTGTTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTC 353
||| | || | | | | | | | | | | | | | | | |
Db 12 CTGCTGTCCCAAGGGACTCCGGGGT CAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTC 71

Qy 354 CTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCATTGCATTTGCTTC 413
||| | | | | | | | | | | | | | |

```

Db          72 CCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGC 131
Qy          414 CTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATC 473
            ||| | ||||| || | | | || | | | || | | | |
Db          132 CTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCATGGGTCTCCAAGTA 191
Qy          474 AACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCAC 533
            ||||| || | ||||| || | | | || | | | || | |
Db          192 AACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGAGCCT---CAC 248
Qy          534 AGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGG 576
            ||| | || ||| ||| ||||| ||||| | | |
Db          249 AGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 291

```

RESULT 42

AX320886

LOCUS AX320886 249 bp DNA linear PAT 14-DEC-2001

DEFINITION Sequence 7 from Patent WO0179272.

ACCESSION AX320886

VERSION AX320886.1 GI:17902435

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Tian,H., Schultz,J. and Shan,B.

TITLE Sitosterolemia susceptibility gene (ssg): compositions and methods
of use

JOURNAL Patent: WO 0179272-A 7 25-OCT-2001;
Tularik Inc. (US)

FEATURES Location/Qualifiers

source 1. .249
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="exon 1 of hSSG"

ORIGIN

Query Match 6.5%; Score 101.6; DB 6; Length 249;
Best Local Similarity 68.4%; Pred. No. 1.2e-19;
Matches 156; Conservative 0; Mismatches 69; Indels 3; Gaps 1;

```

Qy          341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
            || ||| ||||| || | |||| | |||| | || | ||||| ||||| |
Db          25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
Qy          401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
            ||||| |||| | ||||| || | | | || | | | || |
Db          85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCAT 144
Qy          461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCAC 520
            || | || | ||||| || | ||||| ||||| | | | | || | |
Db          145 GGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
Qy          521 AGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAG 568

```

Db 205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAG 249

RESULT 43

AX320883

LOCUS AX320883 2340 bp DNA linear PAT 14-DEC-2001

DEFINITION Sequence 4 from Patent WO0179272.

ACCESSION AX320883

VERSION AX320883.1 GI:17902433

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Tian,H., Schultz,J. and Shan,B.

TITLE Sitosterolemia susceptibility gene (ssg): compositions and methods
of use

JOURNAL Patent: WO 0179272-A 4 25-OCT-2001;
Tularik Inc. (US)

FEATURES Location/Qualifiers

source

1. .2340

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

/note="human sitosterolemia gene (SSG)"

CDS

107. .2062

/note="unnamed protein product; human sitosterolemia
susceptibility gene (SSG) protein"

/codon_start=1

/protein_id="CAD19409.1"

/db_xref="GI:17902434"

/db_xref="REMTREMBL:CAD19409"

/translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
YSVSHRVRPWWDTSCRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGR
LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGN
PGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLDE
PTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGT
PAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS
CHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQ
NLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSD
QESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLA
PHLIGEFLLVLLGIVQNPNI VNSVALLSIAGVLVSGSGLRNIQEMPIPFKIISYFT
FQKYCSEILVNEFYGLNFTCGSSNVSVTNPMCAFTQGIQFIEKTCPGATSRTMNF
LILYSFIPALVILGIVVFKIRDHLISR"

ORIGIN

Query Match 6.5%; Score 101.6; DB 6; Length 2340;

Best Local Similarity 67.4%; Pred. No. 1.3e-19;

Matches 159; Conservative 0; Mismatches 74; Indels 3; Gaps 1;

Qy 341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400

|| ||| ||||| || | |||| | |||| | || ||||| ||||| |

Db 25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84

Qy 401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460

```

      ||||| |||| | ||||| |||| | | | ||| || | |||
Db      85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT 144

Qy      461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGTACAGGGCAC 520
      || | || | ||||| || | ||||| ||||| | | | ||| | | |
Db      145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204

Qy      521 AGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTTACAGCGTCAGGTAAGGGG 576
      ||| || | ||||| | || ||| ||| ||||| ||||| | | |
Db      205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 257

```

RESULT 44

AX685733

LOCUS AX685733 2340 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 5 from Patent WO02081691.

ACCESSION AX685733

VERSION AX685733.1 GI:29371742

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 5 17-OCT-2002;
Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)

FEATURES

source

Location/Qualifiers

1. .2340

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

CDS

107. .2062

/note="unnamed protein product; human ABCG5 (hABCG5)"

/codon_start=1

/protein_id="CAD86572.1"

/db_xref="GI:29371743"

/db_xref="REMTREMBL:CAD86572"

/translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
YSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGR
LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGN
PGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLDE
PTTGLDCMTANQIVVLLVELARRNRIVLTIHQPRSELFQLFDKIAILSFGELIFCGT
PAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS
CHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQ
NLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSD
QESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLA
PHLIGEFLLVLLGIVQNPINVSVALLSIAGVLVSGFRLNIQEMPPIPKIISYFT
FQKYCSEILVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF
LILYSFIPALVILGIVVFKIRDHLISR"

ORIGIN

Query Match 6.5%; Score 101.6; DB 6; Length 2340;
Best Local Similarity 67.4%; Pred. No. 1.3e-19;
Matches 159; Conservative 0; Mismatches 74; Indels 3; Gaps 1;

Qy 341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
 || ||| ||||| || | |||| | |||| | || | ||||| ||||| ||
 Db 25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84

Qy 401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
 ||||| |||| | ||||| |||| | || | || | || | || ||
 Db 85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT 144

Qy 461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCAC 520
 || | || | ||||| || | ||||| |||| | || | || | || |
 Db 145 GGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204

Qy 521 AGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTACAGCGTCAGGTAAGGGG 576
 ||| || ||||| || || |||| |||| | ||||| |||| | || |
 Db 205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 257

RESULT 45

AF320293

LOCUS AF320293 2340 bp mRNA linear PRI 13-DEC-2000

DEFINITION Homo sapiens ABCG5 (ABCG5) mRNA, complete cds.

ACCESSION AF320293

VERSION AF320293.1 GI:11692799

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2340)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J.,
 Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Accumulation of Dietary Cholesterol in Sitosterolemia Caused by
 Mutations in Adjacent ABC Transporters

JOURNAL Science (2001) In press

REFERENCE 2 (bases 1 to 2340)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J.,
 Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Direct Submission

JOURNAL Submitted (09-NOV-2000) Molecular Genetics, University of Texas,
 Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd.,
 Dallas, TX 75390-9046, USA

FEATURES Location/Qualifiers

source 1. .2340
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"

gene 1. .2340
 /gene="ABCG5"

CDS 107. .2062
 /gene="ABCG5"
 /note="ATP-binding cassette, subfamily G, member 5"
 /codon_start=1
 /product="ABCG5"
 /protein_id="AAG40003.1"
 /db_xref="GI:11692800"
 /translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS"

YVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGR
 LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRG
 PGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLDE
 PTTGLDCMTANQIVLLVELARRNRIVLTIHQPRSELFQLFDKIAILSFGELIFCGT
 PAEMLDFNDGCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS
 AI CHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQ
 NLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRAVSD
 QESQDGLYQKWQMMLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLA
 PHLIGEFLLTLLGIVQNPNIIVNSVALLSIAAGVLVSGFGLRNQIEMPIPKIISYFT
 FQKYCSEILVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF
 LILYSFIPALVILGIVVFKIRDHLISR"

ORIGIN

Query Match 6.5%; Score 101.6; DB 9; Length 2340;
 Best Local Similarity 67.4%; Pred. No. 1.3e-19;
 Matches 159; Conservative 0; Mismatches 74; Indels 3; Gaps 1;

Qy	341	CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC	400
Db	25	CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC	84
Qy	401	TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG	460
Db	85	CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT	144
Qy	461	AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCAC	520
Db	145	GGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC	204
Qy	521	AGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGG	576
Db	205	GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG	257

RESULT 46

AX456519

LOCUS AX456519 1920 bp DNA linear PAT 06-JUL-2002

DEFINITION Sequence 41 from Patent WO0227016.

ACCESSION AX456519

VERSION AX456519.1 GI:21715409

KEYWORDS .

SOURCE synthetic construct

ORGANISM synthetic construct
 artificial sequences.

REFERENCE 1

AUTHORS Patel,S.B. and Dean,M.

TITLE Gene involved in dietary sterol absorption and excretion and uses therefor

JOURNAL Patent: WO 0227016-A 41 04-APR-2002;
 THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
 Shailendra B. (US) ; Dean, Michael (US)

FEATURES

source Location/Qualifiers
 1. .1920
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Primer"

```

Query Match          5.9%; Score 93; DB 6; Length 1920;
Best Local Similarity 84.0%; Pred. No. 6.1e-17;
Matches 105; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy      1162 AGCAACCGTGTCGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAA 1221
          ||| |||| || |||| ||||| ||||| ||| |||| |||| ||||| |||||
Db      106 AGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAG 165

Qy      1222 ATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGC 1281
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      166 ATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGC 225

Qy      1282 TCAGG 1286
          |||||
Db      226 TCAGG 230

```

```

LOCUS      AX320887                122 bp      DNA      linear      PAT 14-DEC-2001
DEFINITION Sequence 8 from Patent WO0179272.
ACCESSION  AX320887
VERSION    AX320887.1   GI:17902436
KEYWORDS   .
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
  AUTHORS  Tian,H., Schultz,J. and Shan,B.
  TITLE    Sitosterolemia susceptibility gene (ssg): compositions and methods
            of use
  JOURNAL  Patent: WO 0179272-A 8 25-OCT-2001;
            Tularik Inc. (US)
FEATURES   Location/Qualifiers
     source          1. .122
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="exon 2 of hSSG"

```

```

Query Match          5.7%; Score 90; DB 6; Length 122;
Best Local Similarity 83.6%; Pred. No. 4.7e-16;
Matches 102; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy      1164 CAACCGTGTCTGGGCTTGGTGGAAATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
          | ||| || ||| ||||| ||||| || |||| |||| ||||| |||
Db      1    CCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGAT 60

Qy      1224 CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTC 1283
          ||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db      61    CCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTC 120

Qy      1284 AG 1285

```

Db ||
 121 AG 122

RESULT 48

AF351785/c

LOCUS AF351785 4829 bp mRNA linear ROD 26-AUG-2002

DEFINITION Rattus norvegicus sterolin-2 (Abcg8) mRNA, complete cds.

ACCESSION AF351785

VERSION AF351785.2 GI:22477145

KEYWORDS .

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 4829)

AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H.,
Ose,L., Stalenhoeft,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E.,
Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and
Patel,S.B.

TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterolin-1 and
sterolin-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)

MEDLINE 21344600

PUBMED 11452359

REFERENCE 2 (bases 1 to 4829)

AUTHORS Lu,K., Yu,H., Lee,M. and Patel,S.B.

TITLE Molecular cloning, genomic structure, and characterization of novel
mouse head-to-head tandem ABC transporters

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 4829)

AUTHORS Lu,K., Lee,M. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29407, USA

REFERENCE 4 (bases 1 to 4829)

AUTHORS Lu,K., Yu,H., Lee,M. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-2002) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403, USA

REMARK Sequence update by submitter

COMMENT On Aug 26, 2002 this sequence version replaced gi:15148516.

FEATURES Location/Qualifiers

source 1. .4829
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
gene 1. .4829
 /gene="Abcg8"
CDS 111. .2129
 /gene="Abcg8"
 /codon_start=1


```

/product="sterolin-2"
/protein_id="AAK84831.2"
/db_xref="GI:22477146"
/translation="MAEKTKEETQLWNGTVLQDASSLQDSVFSSES DNSLYFTYSGQS
NTLEVRDLTYQVDMASQVPWFELAQFKLPWRSRGSQDSWDLGIRNLSFKVRSQMLA
IIGSAGCGRATLLDVITGRDHGGMKSGQIWINQGPSTPQLIQKCVAHVRQQDQLLPN
LTVRETLTFIAQMRLPKTFSSQAQRDKRVEDVIAELRLRQCANTRVGNITYVRGVSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNLRVTLRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGVAQHMQYFTS IGYPCPRYSNPADFYVDLTSIDRRSKEQ
EVATMEKARLLAALFLEKVQGFDDFLWKAEAKSLDTGTAVSQTLTQDTNCGTAAELP
GMIQQFTTLIRRIISNDFRDLPTLFHGAELMSLIIGFLYYGHADKPLSFMDMAAL
LFMIGALIPFNVLIDVSKCHSERSLLYELDGLYTAGPYFFAKVLGELPEHCAYVI
IYGMPIYWLNLNLRPGPELFLHFMLLWLNVFCCRTMALAASAMLP TFHMS SFCCNALY
NSFYLTAGFMINLNLWIVPAWISKMSFLRWCFSGMLMQIQFNHGIYTTQIGNLTFVSP
GDAMVTAMDNLNSHPLYAIYLIVIGISCGFLSLYYLSLKFKIKQKSIQDW"

```

ORIGIN

```

Query Match          5.4%;  Score 84;  DB 10;  Length 4829;
Best Local Similarity 77.9%;  Pred. No. 4.1e-14;
Matches 134;  Conservative 0;  Mismatches 20;  Indels 18;  Gaps 2;

Qy      2 GAAGCATCCTGAAGTACAGTCCCATTCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGCC 61
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      172 GAAGCATCCTGGAGTACAGTCCCGTTCACAGCTGGGTCTCCTCTTTGGTCTTCTCAGCC 113

Qy      62 ATGACC-----AGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTT----- 106
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      112 ATGACCTGCGGTGTTGTGCCCTTTGTGTGGCTCCTGAGGCCTCCCCTGCTGTTGGCTAGG 53

Qy      107 ---GGGCTCTCTCTGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCT 155
        || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      52 CCAGGATTCTTTCTGTCTTTGCTCCTTAGAGCTAGGGCACTTGAGTCCTCCT 1

```

RESULT 49

AC146282/c

LOCUS AC146282 135280 bp DNA linear HTG 02-AUG-2003
 DEFINITION Takifugu rubripes clone MRC-186C24, WORKING DRAFT SEQUENCE, 7
 unordered pieces.

ACCESSION AC146282

VERSION AC146282.1 GI:33413347

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Takifugu rubripes (Fugu rubripes)

ORGANISM Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontoidea; Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 135280)

AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
 Peng,Z., Malinov,I. and Rubin,E.M.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 135280)

AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
 Peng,Z., Malinov,I. and Rubin,E.M.

TITLE Direct Submission

Qy 1278 CAGCTCAGGTAAGTGCCT 1295
|| ||||| | | |
Db 33269 CAACTCAGGTTTGCACGT 33252

RESULT 50
AX685731/c

LOCUS AX685731 2019 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 3 from Patent WO02081691.

ACCESSION AX685731

VERSION AX685731.1 GI:29371740

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Hobbs, H.H., Shan, B., Barnes, R. and Tian, H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 3 17-OCT-2002;
Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)

FEATURES Location/Qualifiers

source

1. .2019
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"

CDS

1. .2019
/note="unnamed protein product; mouse ABCG8 (mABCG8)"
/codon_start=1
/protein_id="CAD86571.1"
/db_xref="GI:29371741"
/db_xref="REMTREMBL:CAD86571"
/translation="MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQS
NTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLA
IIGSSGCGRASLLDVITGRGHGGMKSGQI WINGQPSTPQLVRKCVAVHRQHDQLLPN
LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGN TYVRGVSGGER
RRVSIGVQLLWNP GILILDEPTSGLD SFTAHNLVTTLSRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKER
EVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTLTQD TDCGTAVELP
GMIEQFSTLIRROI SNDFRDLPTLLIHGSEACLMSLIIGFLYYGHGAKQLS FMDTAAL
LFMIGALIPFNVILDV VSKCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVI
IYAMPIYWLTNLRPVPELFLHFLLVLVVFCRTMALAASAMLP TFHMSSF CFCNALY
NSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSIL
GDTMISAMDLNSHPLYAIYLIVIGISYGF LFLYYLSLKLKQKSIQDW"

ORIGIN

Query Match 4.0%; Score 63; DB 6; Length 2019;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTCTCAGC 60

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 63 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTCTCAGC 4

Qy 61 CAT 63

|||

Db

3 CAT 1

Search completed: April 29, 2004, 17:06:37
Job time : 6741.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 14:53:09 ; Search time 756.83 Seconds
(without alignments)
8812.639 Million cell updates/sec

Title: US-09-989-981A-9_COPY_3436_5005
Perfect score: 1570
Sequence: 1 cgaagcatcctgaagtacag.....ctagagagcaaaccagagc 1570

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	1568	99.9	6043	7	AAD48884	Aad48884 ABCG5-ABC
2	358.6	22.8	359	7	AAD48885	Aad48885 Control D
3	284.4	18.1	2354	6	ABK51685	Abk51685 Mouse ABC
4	242.8	15.5	5460	6	ABK51683	Abk51683 Human ABC
5	215	13.7	2512	9	ADB62671	Adb62671 Human cDN
6	191.4	12.2	2258	6	AAD22008	Aad22008 Mouse sit
7	151.2	9.6	226	3	AAA10131	Aaa10131 Rat liver

	8	151.2	9.6	226	6	AAD42155	Aad42155	Rat targe
	9	150.8	9.6	235	3	AAA87503	Aaa87503	Rat hepat
	10	146.4	9.3	1915	6	ABK51684	Abk51684	DNA encod
	11	146.4	9.3	1959	7	AAD48880	Aad48880	Mouse ABC
	12	135.8	8.6	2035	6	ABK51686	Abk51686	cDNA enco
	13	107	6.8	2516	6	ABK51682	Abk51682	Human ABC
	14	101.6	6.5	249	6	AAD22010	Aad22010	Human sit
	15	101.6	6.5	2340	6	AAD22009	Aad22009	Human sit
	16	101.6	6.5	2340	7	AAD48882	Aad48882	Human ABC
c	17	97	6.2	2564	6	ABN90022	Abn90022	Mouse clo
	18	93	5.9	1920	6	ABK51681	Abk51681	DNA encod
	19	90	5.7	122	6	AAD22011	Aad22011	Human sit
c	20	63	4.0	2019	7	AAD48881	Aad48881	Mouse ABC
	21	44.6	2.8	2000	7	ADA71938	Ada71938	Rice gene
c	22	38.2	2.4	2000	7	ADA71938	Ada71938	Rice gene
c	23	37	2.4	2692	9	ADE57382	Ade57382	Rat gene
c	24	37	2.4	4590	5	AAH24065	Aah24065	Yeast AOD
	25	36.8	2.3	6843	6	ABN83968	Abn83968	Human gen
c	26	36.2	2.3	541	5	AAS93276	Aas93276	DNA encod
c	27	36.2	2.3	2236	8	ADA08012	Ada08012	cDNA enco
c	28	36.2	2.3	2488	8	ADA08010	Ada08010	cDNA enco
c	29	36.2	2.3	4003	5	AAS93277	Aas93277	DNA encod
	30	35.8	2.3	397	4	AAL19738	Aal19738	Human bre
c	31	35.8	2.3	493	4	AAH13437	Aah13437	Human cDN
c	32	35.8	2.3	639	6	ABN73379	Abn73379	Bovine em
	33	35.6	2.3	257	6	ABL79659	Abl79659	Human ova
c	34	35.6	2.3	313	2	AAQ60875	Aaq60875	Human bra
c	35	35.6	2.3	330	6	ABN94395	Abn94395	Gene #893
c	36	35.6	2.3	440	6	ABL87579	Abl87579	Human ova
c	37	35.6	2.3	1333	3	AAC76886	Aac76886	Human ORF
	38	35.6	2.3	2474	4	AAH18291	Aah18291	Human cDN
	39	35.6	2.3	21632	4	AAS42019	Aas42019	Genomic s
c	40	35.6	2.3	122186	4	AAC89560	Aac89560	Human his
c	41	35.4	2.3	639	6	ABN73289	Abn73289	Bovine em
	42	35.2	2.2	414	4	AAL18444	Aal18444	Human bre
	43	35.2	2.2	416	4	AAL10158	Aal10158	Human bre
	44	35.2	2.2	459	4	AAL09772	Aal09772	Human bre
c	45	35.2	2.2	760	2	AAZ16007	Aaz16007	Human gen
c	46	35.2	2.2	2412	5	AAS68011	Aas68011	DNA encod
c	47	35.2	2.2	2412	9	ADC32278	Adc32278	Human nov
c	48	35.2	2.2	4866	6	ABN59902	Abn59902	Novel hum
c	49	35.2	2.2	5011	5	AAS72352	Aas72352	DNA encod
c	50	35	2.2	907	6	ABK33144	Abk33144	DNA encod

ALIGNMENTS

RESULT 1

AAD48884

ID AAD48884 standard; DNA; 6043 BP.

XX

AC AAD48884;

XX

DT 24-MAR-2003 (first entry)

XX

DE ABCG5-ABCG8 DNA.

XX
 KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW ATP-binding cassette; sitosterolaemia susceptibility gene; SSG; ABCG5;
 KW ds.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT exon complement(3. .104)
 FT /*tag= a
 FT /number= 2
 FT /note= "Corresponds to ABCG8 gene"
 FT intron complement(105. .3435)
 FT /*tag= b
 FT /number= 1
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT /note= "Corresponds to ABCG8 gene"
 FT misc_feature complement(1098. .1377)
 FT /*tag= c
 FT /note= "ABCG8 intron1 conserved region"
 FT misc_feature complement(3250. .3294)
 FT /*tag= d
 FT /note= "ABCG8 intron1 conserved region"
 FT exon 3436. .3498
 FT /*tag= e
 FT /number= 1
 FT /note= "Corresponds to ABCG8 gene"
 FT exon 3858. .4003
 FT /*tag= f
 FT /number= 1
 FT /note= "Corresponds to ABCG5 gene"
 FT intron 4004. .4598
 FT /*tag= g
 FT /number= 1
 FT /note= "Corresponds to ABCG5 gene"
 FT exon 4599. .4720
 FT /*tag= h
 FT /number= 2
 FT /note= "Corresponds to ABCG5 gene"
 FT intron 4721. .6043
 FT /*tag= i
 FT /number= 2
 FT /partial
 FT /note= "Corresponds to ABCG5 gene"
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
PT related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies.

XX

PS Disclosure; Fig 3; 94pp; English.

XX

CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is ABCG8- ABCG5 DNA

XX

SQ Sequence 6043 BP; 1378 A; 1509 C; 1497 G; 1654 T; 0 U; 5 Other;

Query Match 99.9%; Score 1568; DB 7; Length 6043;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC	60
Db	3436	CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC	3495
Qy	61	CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC	120
Db	3496	CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC	3555
Qy	121	TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTCTCCCAGCATT	180
Db	3556	TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTCTCCCAGCATT	3615
Qy	181	CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG	240
Db	3616	CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG	3675
Qy	241	ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTAT	300
Db	3676	ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTAT	3735
Qy	301	CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT	360
Db	3736	CTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT	3795
Qy	361	AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCATTGCTTCCTGCTAG	420
Db	3796	AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCATTGCTTCCTGCTAG	3855
Qy	421	CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG	480

Db	3856	 CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG	3915
Qy	481	GGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAG	540
Db	3916	 GGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAG	3975
Qy	541	GTGTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCT	600
Db	3976	 GTGTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCT	4035
Qy	601	CTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGC	660
Db	4036	 CTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGC	4095
Qy	661	AGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTG	720
Db	4096	 AGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTG	4155
Qy	721	CGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCAC	780
Db	4156	 CGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCAC	4215
Qy	781	TGATTTCTGCTCTCCCCCTTCCTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGG	840
Db	4216	 TGATTTCTGCTCTCCCCCTTCCTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGG	4275
Qy	841	CTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGG	900
Db	4276	 CTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGG	4335
Qy	901	GTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAA	960
Db	4336	 GTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAA	4395
Qy	961	TCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCC	1020
Db	4396	 TCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCC	4455
Qy	1021	AAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACT	1080
Db	4456	 AAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACT	4515
Qy	1081	GGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTG	1140
Db	4516	 GGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTG	4575
Qy	1141	CCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGCCTTGGTGGAACATCAAATCATGC	1200
Db	4576	 CCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGCCTTGGTGGAACATCAAATCATGC	4635
Qy	1201	CAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAG	1260
Db	4636	 CAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAG	4695
Qy	1261	ATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCCTGTACTTC	1320

Db 4696 ATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCCTGTACTTC 4755
 Qy 1321 TAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAATGTTTAAGAAGTRAGTTTA 1380
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 4756 TAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAATGTTTAAGAAGTRAGTTTA 4815
 Qy 1381 AGTTGTAGAGAGGCAGCCATGCATTTGGCATTTGAATACAATCTGGTGACTTGTCTGGCT 1440
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 4816 AGTTGTAGAGAGGCAGCCATGCATTTGGCATTTGAATACAATCTGGTGACTTGTCTGGCT 4875
 Qy 1441 GCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAG 1500
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 4876 GCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAG 4935
 Qy 1501 TCCTGCCTAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCA 1560
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 4936 TCCTGCCTAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCA 4995
 Qy 1561 AACCCAGAGC 1570
 |||||||||
 Db 4996 AACCCAGAGC 5005

RESULT 2

AAD48885

ID AAD48885 standard; DNA; 359 BP.

XX

AC AAD48885;

XX

DT 24-MAR-2003 (first entry)

XX

DE Control DNA fragment flanked by ABCG5-ABCG8 DNA sequence.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW ATP-binding cassette; sitosterolaemia susceptibility gene; SSG; ABCG5;

KW ds.

XX

OS Unidentified.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
PT related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies.
XX
PS Disclosure; Fig 3; 94pp; English.
XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is control DNA fragment
CC flanked by ABCG5-ABCG8 DNA sequence
XX
SQ Sequence 359 BP; 68 A; 103 C; 87 G; 100 T; 0 U; 1 Other;

Query Match 22.8%; Score 358.6; DB 7; Length 359;
Best Local Similarity 100.0%; Pred. No. 7.5e-103;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	64	GACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCTGCTGTTGGGCTCTCTCTGTCTTT	123
Db	1	GACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCTGCTGTTGGGCTCTCTCTGTCTTT	60
Qy	124	GCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTTCCT	183
Db	61	GCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTTCCT	120
Qy	184	YTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTGAGATAAGGACA	243
Db	121	YTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTGAGATAAGGACA	180
Qy	244	CTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTC	303
Db	181	CTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTC	240
Qy	304	ACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTTAAA	363
Db	241	ACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTTAAA	300
Qy	364	GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAGCC	422
Db	301	GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAGCC	359

RESULT 3

ABK51685

ID ABK51685 standard; cDNA; 2354 BP.

XX

AC ABK51685;

XX

DT 30-JUL-2002 (first entry)

XX

DE Mouse ABCG5 cDNA sequence.

XX
 KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ss.
 XX
 OS Mus sp.
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Example 3; Page 45; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of the
 CC invention
 XX
 SQ Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 U; 0 Other;

 Query Match 18.1%; Score 284.4; DB 6; Length 2354;
 Best Local Similarity 98.0%; Pred. No. 8.1e-79;
 Matches 288; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Qy      285 ATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCC 344
      |||
Db      1  ATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCC 60

Qy      345 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACTTGC 404
      |||
Db      61 CATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACTTGC 120

Qy      405 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 464
      |||
Db      121 ATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGG 180

Qy      465 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 524
      |||
Db      181 CCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAG 240

Qy      525 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGAC 578
      |||
Db      241 GCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTC 294

```

RESULT 4

ABK51683

ID ABK51683 standard; DNA; 5460 BP.

XX

AC ABK51683;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 upstream genomic sequence, exon 1, intron 1 and exon 2.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW chromosome 2p21; ds.

XX

OS Homo sapiens.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.

XX

XX
 AC ADB62671;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cDNA encoding clone LIVER20030650.
 XX
 KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 KW tissue regeneration; cell regeneration; membrane protein;
 KW signal transduction-related protein; transcription-related protein;
 KW osteoporosis; neurological disease; cancer; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1469..2239
 FT /*tag= a
 FT /product= "Clone LIVER20030650 protein"
 XX
 PN EP1308459-A2.
 XX
 PD 07-MAY-2003.
 XX
 PF 28-MAR-2002; 2002EP-00007401.
 XX
 PR 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI; 2003-450961/43.
 DR P-PSDB; ADB64641.
 XX
 PT New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 XX
 PS Claim 1; Page; 222pp; English.
 XX
 CC The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesising the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or

CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.

XX

SQ Sequence 2512 BP; 543 A; 675 C; 701 G; 593 T; 0 U; 0 Other;

Query Match 13.7%; Score 215; DB 9; Length 2512;

Best Local Similarity 54.5%; Pred. No. 9.2e-57;

Matches 576; Conservative 0; Mismatches 450; Indels 31; Gaps 6;

```
Qy      237 AAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAACCTG 296
      ||||| | ||||||||||||||||||||||||| || ||||| | |||
Db      1 AAGGACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCATTCCTG 60

Qy      297 TTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTC 356
      | ||| | | | | | | | | | | | | | | | | | | | | |
Db      61 CTGTCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCC 120

Qy      357 AGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCATTGCAATTTGCTTCCTG 416
      | | |||| | |||| | || | ||||| ||||| | | ||||| |||
Db      121 AACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTG 180

Qy      417 CTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCTCACATCAAC 476
      | ||||||||| || | | | || | || | || | || | || | |||
Db      181 TTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAAC 240

Qy      477 AGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGC 536
      ||||| || | ||||||||||||| | | | || | | | | || | |||||
Db      241 AGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGC 297

Qy      537 TTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTA 596
      | || |||| |||| ||||||||||||||||||| || |||
Db      298 CTGGGCATCCTCCATGCCTCCTACAGCGTCAGGTAAGGCAGAGCCC-----TTGCTG 349

Qy      597 GGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTC 656
      || || || || || || || || || || || || || || || || ||
Db      350 CTGCTGCTCCCCCAGGAGTGCGGGGCCGGCGCTCACCCCTCTGCTGCCTTTCTTCACTC 409

Qy      657 CAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCT 716
      || | | | || || | || || || || || || || || || || ||
Db      410 TTTAAGTGCCAGTCTGGGCACTTCGGGCTCCCTCTTAGTGGATCGGGTGGAGAGAGGAG 469

Qy      717 CCTGCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTC 776
      | | ||||| | || || | | | | | | | | | | | | | |||
Db      470 AGGGAGAAGGGCTGTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCCTCCCCCTGCCCA 529

Qy      777 TCACTGATTTCTGCTCTCCCTTCCTTGACTC-GCCCACCACCTGTCTGTGTAGATGGA 835
      | | | | | | | | | | | | | | | | | | | |||
Db      530 GGGAAGGGCCTGGGCATAAAACAAAGTGGCAGCAGTGCCTGCCAACCAGTGTCTACGGC 589
```

Qy	836	GAAGGCTC-----GGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGA	889
Db	590	CTGCCCTCTGTGGATGGGAATGGGGTACTGCGAATGCAAGGAGTCTTGAAACCTGGTGA	649
Qy	890	AGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCCAGTCAG	949
Db	650	AAGAATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTTCAGAGCAAC	703
Qy	950	AGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACCTTGTC	1009
Db	704	TCCAGCACAGGCCCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTTCT-	762
Qy	1010	ATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTC	1069
Db	763	-----ACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCCTGTCCGGAGACTACTG	816
Qy	1070	CAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTA	1129
Db	817	GGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCCCTTCCAGGGCCCCA	876
Qy	1130	CTCTGCCTTTGCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTGCGGGCCTTGGTGGAACA	1189
Db	877	AGCTGCCTTTGCCCTTTCCTGGGGTTTCCTTTAAAGCCACCGGTGAGGCCCTGGTGGGACA	936
Qy	1190	TCAAATCATGCCAGCAGAAGTGGGACAGGCCAAATCCTCAAAGATGTCTCCTTGTACATCG	1249
Db	937	TCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGG	996
Qy	1250	AGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGG	1286
Db	997	AGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGG	1033

RESULT 6

AAD22008

ID AAD22008 standard; DNA; 2258 BP.

XX

AC AAD22008;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG).

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT CDS 47. .2005

FT /*tag= a

FT /product= "Mouse SSG protein"

XX

PN WO200179272-A2.

XX

RESULT 7

AAA10131

ID AAA10131 standard; cDNA; 226 BP.

XX

AC AAA10131;

XX

DT 03-JUL-2000 (first entry)

XX

DE Rat liver toxicological response marker, SEQ ID NO:24.

XX

KW Toxicological response marker; rat; liver; expression pattern;

KW toxicity screening; toxic compound; polycyclic aromatic hydrocarbon; PAH;

KW benzo(a)pyrene; clofibrate; acetaminophen; ss.

XX

OS Rattus norvegicus.

XX

PN WO200012760-A2.

XX

PD 09-MAR-2000.

XX

PF 27-AUG-1999; 99WO-US019768.

XX

PR 28-AUG-1998; 98US-00141825.

PR 13-OCT-1998; 98US-00172108.

PR 13-OCT-1998; 98US-00172711.

XX

PA (INCY-) INCYTE PHARM INC.

XX

PI Cunningham MJ, Zweiger GB, Panzer SR, Seilhamer JJ;

XX

DR WPI; 2000-237888/20.

XX

PT Isolated and purified nucleic acid molecules used as toxicological
PT response markers for detecting and diagnosing a potential toxicological
PT response in a mammalian subject to a test compound or molecule.

XX

PS Claim 6; Page 44; 76pp; English.

XX

CC Sequences AAA10108-A10224 represent rat liver toxicological response
CC markers. These were identified by their pattern of at least twofold
CC upregulation or downregulation of expression in rat liver treated with a
CC toxic compound (e.g., clofibrate, acetaminophen or polycyclic aromatic
CC hydrocarbons (PAHs) such as benzo(a)pyrene). Fluorescently labelled rat
CC liver mRNA was contacted with a microarray comprising a library of rat
CC cDNA molecules. Twofold or larger changes in hybridisation were only
CC observed between the sample mRNA and sequences AAA10108-A10224. In
CC particular, sequences AAA10110, AAA10116, AAA10117, AAA10120, AAA10126,
CC AAA10133, AAA10138, AAA10140, AAA10142-A10144, AAA10146, AAA10149,
CC AAA10164, AAA10174, AAA10185, AAA10188, AAA10189, AAA10201 and AAA10205
CC were all upregulated in samples treated with known toxic compounds
CC relative to untreated samples, while sequences AAA10150, AAA10156,
CC AAA10157, AAA10159-A10163, AAA10166, AAA10168, AAA10170, AAA10175,
CC AAA10178, AAA10181, AAA10192, AAA10194, AAA10197, AAA10202, AAA10209,
CC AAA10210, AAA10212 and AAA10222 were all downregulated. Expression of


```

XX
PR 04-MAY-1998; 98US-0084029P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Cunningham MJ, Zweiger GB, Panzer SR, Seilhamer JJ;
XX
DR WPI; 2002-536048/57.
XX
PT Composition useful as hybridizable array elements in a microarray, for
PT screening compounds for toxicological responses, has many polynucleotide
PT targets derived from rat liver cDNA libraries and rat kidney libraries.
XX
PS Claim 1; Col 29-30; 23pp; English.
XX
CC The invention relates to a composition comprising a plurality of
CC polynucleotide targets. The polynucleotide targets are derived from rat
CC liver cDNA libraries and rat kidney libraries. The composition can be
CC immobilised on a substrate and used as hybridisable array elements in a
CC microarray format. The microarray is used to characterise gene expression
CC patterns associated with novel compounds to elucidate any toxicological
CC effects or to monitor the effects of therapeutic treatments, where
CC toxicological effects may be expected. The composition is also useful for
CC screening compounds and/or therapeutic treatments for potential
CC toxicological effects and for screening a sample's toxicological response
CC to a particular test compound. The present sequence is rat target DNA
XX
SQ Sequence 226 BP; 51 A; 62 C; 67 G; 46 T; 0 U; 0 Other;

```

```

RESULT 9
AAA87503
ID    AAA87503 standard; DNA; 235 BP.
XX
AC    AAA87503;
XX
DT    08-JAN-2001 (first entry)

```


XX
 AC ABK51684;
 XX
 DT 30-JUL-2002 (first entry)
 XX
 DE DNA encoding mouse ABCG5 protein.
 XX
 KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ds.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1. .1915
 FT /*tag= a
 FT /partial
 FT /product= "Mouse ABCG5 protein"
 FT /transl_except= (pos: 1912. .1915, aa: LGIVIFKVRDY LISR)
 FT /note= "This sequence lacks a stop codon"
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 DR P-PSDB; AAU96985.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Example 3; Page 42-43; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the

PR 28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

DR P-PSDB; AAE31702.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-

PT related disorders e.g. sitosterolemia, hypercholesterolemia,

PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or

PT nutritional deficiencies.

XX

PS Claim 11; Page 73; 94pp; English.

XX

CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is mouse ABCG5 DNA

XX

SQ Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 U; 0 Other;

Query Match 9.3%; Score 146.4; DB 7; Length 1959;

Best Local Similarity 96.2%; Pred. No. 4.8e-35;

Matches 150; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 423 ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 482
|||||

Db 1 ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 60

Qy 483 TCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 542
|||||

Db 61 TCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120

Qy 543 GTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGAC 578
|||||

Db 121 GTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTGTC 156

RESULT 12

ABK51686

ID ABK51686 standard; cDNA; 2035 BP.

XX

AC ABK51686;

XX

DT 07-AUG-2003 (revised)

DT 30-JUL-2002 (first entry)

XX

DE cDNA encoding rat ABCG5 protein.

XX

KW Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; ss;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 XX
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 8. .1965
 FT /*tag= a
 FT /product= "Rat ABCG5 protein"
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 DR P-PSDB; AAU96986.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Example 3; Page 45-46; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the rat ABCG5 protein of the invention. (Updated on
 CC 07-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 2035 BP; 481 A; 533 C; 537 G; 484 T; 0 U; 0 Other;

Query Match

8.6%; Score 135.8; DB 6; Length 2035;

Best Local Similarity 89.6%; Pred. No. 1.1e-31;
Matches 146; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy      416 GCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAA 475
          ||| |||||
Db      1  GCTGGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACAAACA 60

Qy      476 CAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAG 535
          |||||
Db      61 CAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTCAGAGGCTCGGCACAG 120

Qy      536 CTTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGAC 578
          |||||
Db      121 CTTAGGTGTCCTGAATGTGTCTTCAGCGTCAGCAACCGTGTG 163
```

RESULT 13

ABK51682

ID ABK51682 standard; cDNA; 2516 BP.

XX

AC ABK51682;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 cDNA sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
chromosome 2p21; ss.

XX

OS Homo sapiens.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
acid encoding the polypeptide, useful for treating sitosterolemia,
arteriosclerosis and heart diseases.

XX

PS Example 3; Page 37-38; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
predisposition for developing sitosterolemia, arteriosclerosis or heart

CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the cDNA sequence of human ABCG5 gene located on
 CC chromosome 2p21

XX

SQ Sequence 2516 BP; 601 A; 631 C; 636 G; 648 T; 0 U; 0 Other;

Query Match 6.8%; Score 107; DB 6; Length 2516;
 Best Local Similarity 63.6%; Pred. No. 1.8e-22;
 Matches 180; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

QY 294 CTGTTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTC 353
 ||| | || || | | | | || | || | | ||| |||
 Db 12 CTGCTGTCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTC 71
 QY 354 CTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAGTTGCATTTGCTTC 413
 | || | |||| | |||| | || | ||||| ||||| | ||||| |
 Db 72 CCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGC 131
 QY 414 CTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATC 473
 ||| | ||||| ||||| || | | || | || | || | || | || |
 Db 132 CTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTA 191
 QY 474 AACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCAC 533
 ||||| || | ||||| ||||| | | | || | | | || | || |
 Db 192 AACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGAGCCT---CAC 248
 QY 534 AGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGG 576
 ||| | || ||| ||| ||||| ||||| || | |
 Db 249 AGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 291

RESULT 14

AAD22010

ID AAD22010 standard; DNA; 249 BP.

XX

AC AAD22010;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG) exon 1.

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
KW xanthoma; haemolytic anaemia; transgenic animal; ds.
XX
OS Homo sapiens.
XX
PN WO200179272-A2.
XX
PD 25-OCT-2001.
XX
PF 18-APR-2001; 2001WO-US012758.
XX
PR 18-APR-2000; 2000US-0198465P.
PR 15-MAY-2000; 2000US-0204234P.
XX
PA (TULA-) TULARIK INC.
XX
PI Tian H, Schultz J, Shan B;
XX
DR WPI; 2002-017598/02.
XX
PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
PT useful for screening a compound that increases the level of expression or
PT activity of SSG polypeptide for treating sterol-related disorder.
XX
PS Claim 73; Fig 14B; 105pp; English.
XX
CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
CC identifying a compound useful in the treatment or prevention of a sterol-
CC related disorder, including sitosterolaemia, hyperlipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolaemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolaemia and other
CC cardiovascular disorders, for forensics and paternity determinations, and
CC for treating any of a large number of SSG associated diseases. The
CC present sequence is an exon of human SSG DNA
XX
SQ Sequence 249 BP; 44 A; 86 C; 74 G; 45 T; 0 U; 0 Other;

Query Match 6.5%; Score 101.6; DB 6; Length 249;
Best Local Similarity 68.4%; Pred. No. 2.4e-21;
Matches 156; Conservative 0; Mismatches 69; Indels 3; Gaps 1;

Qy 341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
|| ||| ||||| || | |||| | |||| | || | ||||| ||||| |
Db 25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
Qy 401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
||||| |||| | ||||| |||| | | | ||| || | || |||
Db 85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCAT 144

Qy 461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCAC 520
 || | || | ||||| || | ||||| || | || | || | || |
 Db 145 GGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
 Qy 521 AGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAG 568
 ||| || ||||| || |||| |||| ||||| |||||
 Db 205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAG 249

RESULT 15

AAD22009

ID AAD22009 standard; DNA; 2340 BP.

XX

AC AAD22009;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG).

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 107..2062

FT /*tag= a

FT /product= "Human SSG protein"

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX

DR WPI; 2002-017598/02.

DR P-PSDB; AAE13290.

XX

PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
 PT useful for screening a compound that increases the level of expression or
 PT activity of SSG polypeptide for treating sterol-related disorder.

XX

PS Claim 8; Fig 8; 105pp; English.

XX

CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for

CC identifying a compound useful in the treatment or prevention of a sterol-
 CC related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations, and
 CC for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG DNA. Human SSG is located on chromosome
 CC 2p21

XX

SQ Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

Query Match 6.5%; Score 101.6; DB 6; Length 2340;
 Best Local Similarity 67.4%; Pred. No. 8.9e-21;
 Matches 159; Conservative 0; Mismatches 74; Indels 3; Gaps 1;

Qy 341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
 || ||| ||||| || | |||| | |||| | || | ||||| ||||| |||
 Db 25 CTGCCACGGGCTCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
 Qy 401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
 ||||| |||| | ||||| |||| | || | ||| || | || |||
 Db 85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCAT 144
 Qy 461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCAC 520
 || | || | ||||| || | ||||| |||| | || | ||| || |
 Db 145 GGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
 Qy 521 AGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCCTACAGCGTCAGGTAAGGGG 576
 ||| || ||||| | || |||| ||| ||||| ||||| || | ||
 Db 205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 257

RESULT 16

AAD48882

ID AAD48882 standard; DNA; 2340 BP.

XX

AC AAD48882;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG5 DNA.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 107. .2062
 FT /*tag= a
 FT /product= "hABCG5 protein"
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR P-PSDB; AAE31704.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 11; Page 77; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is human ABCG5 DNA
 XX
 SQ Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

Query Match 6.5%; Score 101.6; DB 7; Length 2340;
 Best Local Similarity 67.4%; Pred. No. 8.9e-21;
 Matches 159; Conservative 0; Mismatches 74; Indels 3; Gaps 1;

Qy 341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
 || ||| ||||| || | |||| | |||| | || | ||||| ||||| |
 Db 25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
 Qy 401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
 ||||| |||| | ||||| |||| | | | ||| || | || |||
 Db 85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT 144
 Qy 461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGTCACGGGCAC 520
 || | || | ||||| || | ||||| ||||| | | | ||| | | |
 Db 145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
 Qy 521 AGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGG 576
 ||| || ||||| | || |||| |||| | ||||| ||||| | | |

Db 205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 257

RESULT 17

ABN90022/c

ID ABN90022 standard; cDNA; 2564 BP.

XX

AC ABN90022;

XX

DT 16-AUG-2002 (first entry)

XX

DE Mouse clone IMX3_67 extended sequence.

XX

KW Mouse; antiinflammatory; gene therapy; ileitis; DST; ss; TOGA;

KW digital sequence tag; total gene expression analysis.

XX

OS Mus musculus.

XX

PN WO200231114-A2.

XX

PD 18-APR-2002.

XX

PF 11-OCT-2001; 2001WO-US032091.

XX

PR 11-OCT-2000; 2000US-0239483P.

XX

PA (DIGI-) DIGITAL GENE TECHNOLOGIES INC.

XX

PI Viney JL, Sims JE, Dubose RF, Baum PR, Hasel KW, Hilbush BS;

XX

DR WPI; 2002-426279/45.

XX

PT New isolated nucleic acid molecules that are associated with ileitis, for
PT preventing, treating, modulating and diagnosing ileitis in a mammalian
PT subject.

XX

PS Claim 1; Page 266-268; 273pp; English.

XX

CC The invention relates to a novel isolated nucleic acid molecule
CC comprising a polynucleotide having one of 90 polynucleotide sequences,
CC given in the specification. The polynucleotides of the invention have
CC antiinflammatory activity, and may have a use in gene therapy. The
CC polynucleotide or a polypeptide encoded by it is used for preventing,
CC treating, modulating or ameliorating a medical condition such as ileitis.
CC The polypeptide or polynucleotide is also useful for manufacturing a
CC medicament for treating ileitis. The sequence represents a an extended
CC cDNA digital sequence tag obtained from a mouse clone by the TOGA (total
CC gene expression analysis) method

XX

SQ Sequence 2564 BP; 623 A; 722 C; 638 G; 581 T; 0 U; 0 Other;

Query Match 6.2%; Score 97; DB 6; Length 2564;

Best Local Similarity 100.0%; Pred. No. 2.7e-19;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60

|||||

Db 97 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 38

Qy 61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCC 97

Db 37 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCC 1

RESULT 18

ABK51681

ID ABK51681 standard; DNA; 1920 BP.

XX

AC ABK51681;

XX

DT 30-JUL-2002 (first entry)

XX

DE DNA encoding human ABCG5 protein.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW chromosome 2p21; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1. .1920

FT /*tag= a

FT /product= "Human ABCG5 protein"

FT /transl_except= (pos: 4. .9, aa: GDLSSLTPGGSMGL)

FT /note= "This sequence contains 13 exons"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR P-PSDB; AAU98984.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Claim 38; Page 36-37; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette

CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a

CC predisposition for developing sitosterolemia, arteriosclerosis or heart

CC disease. The molecules of the invention are also useful for identifying a

CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the human ABCG5 gene located on chromosome 2p21.
 CC This sequence encodes the human ABCG5 protein of the invention
 XX
 SQ Sequence 1920 BP; 440 A; 503 C; 486 G; 491 T; 0 U; 0 Other;

Query Match 5.9%; Score 93; DB 6; Length 1920;
 Best Local Similarity 84.0%; Pred. No. 4.3e-18;
 Matches 105; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1162 AGCAACCGTGTCGGGCCTTGGTGGGAACATCAAATCATGCCAGCAGAAAGTGGGACAGGCAA 1221
 ||| |||| || |||| ||||| ||||| ||| ||| |||| |||||
 Db 106 AGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAG 165
 QY 1222 ATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGC 1281
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 166 ATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGC 225
 QY 1282 TCAGG 1286
 |||||
 Db 226 TCAGG 230

RESULT 19

AAD22011

ID AAD22011 standard; DNA; 122 BP.

XX

AC AAD22011;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG) exon 2.

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; ds.

XX

OS Homo sapiens.

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

```

XX 18-APR-2001; 2001WO-US012758.
XX
XX 18-APR-2000; 2000US-0198465P.
XX 15-MAY-2000; 2000US-0204234P.
XX
XX (TULA-) TULARIK INC.
XX
XX Tian H, Schultz J, Shan B;
XX
XX WPI; 2002-017598/02.
XX
XX Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
XX useful for screening a compound that increases the level of expression or
XX activity of SSG polypeptide for treating sterol-related disorder.
XX
XX Claim 73; Fig 14B; 105pp; English.
XX
XX The invention relates to an isolated Sitosterolaemia Susceptibility Gene
XX (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
XX binding cassette (ABC) family cholesterol transporter. SSG is useful for
XX identifying a compound useful in the treatment or prevention of a sterol-
XX related disorder, including sitosterolaemia, hyperlipidaemia,
XX hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
XX nutritional deficiencies. SSG is also useful for treating cholesterol-
XX associated diseases or conditions including coronary heart disease and
XX other cardiovascular diseases, and sitosterolaemia-associated condition
XX including arthritis, xanthomas and chronic haemolytic anaemia. SSG
XX expression cassette is useful in the production of transgenic non-human
XX animals. SSG genes and their homologues are useful as tools for a number
XX of applications including diagnosing sitosterolaemia and other
XX cardiovascular disorders, for forensics and paternity determinations, and
XX for treating any of a large number of SSG associated diseases. The
XX present sequence is an exon of human SSG DNA
XX
XX Sequence 122 BP; 27 A; 34 C; 38 G; 23 T; 0 U; 0 Other;
XX
XX Query Match 5.7%; Score 90; DB 6; Length 122;
XX Best Local Similarity 83.6%; Pred. No. 7.7e-18;
XX Matches 102; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
XX
XX QY 1164 CAACCGTGTGCGGCCCTTGTTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 1 CCACCGCGTGAGGCCCTTGTTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGAT 60
XX
XX QY 1224 CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTC 1283
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 61 CCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTC 120
XX
XX QY 1284 AG 1285
XX | |
XX Db 121 AG 122
XX
XX RESULT 20
XX AAD48881/c
XX ID AAD48881 standard; DNA; 2019 BP.

```

XX
 AC AAD48881;
 XX
 DT 24-MAR-2003 (first entry)
 XX
 DE Mouse ABCG8 DNA.
 XX
 KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1. .2019
 FT /*tag= a
 FT /product= "mABCG8 protein"
 FT /transl_except= (pos:1318. .1320, aa:Leu)
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR P-PSDB; AAE31703.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 13; Page 75; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is mouse ABCG8 DNA
 XX
 SQ Sequence 2019 BP; 444 A; 598 C; 510 G; 467 T; 0 U; 0 Other;

Query Match

4.0%; Score 63; DB 7; Length 2019;

Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
|||||
Db 63 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 4

Qy 61 CAT 63
|||
Db 3 CAT 1

RESULT 21

ADA71938

ID ADA71938 standard; DNA; 2000 BP.

XX

AC ADA71938;

XX

DT 20-NOV-2003 (first entry)

XX

DE Rice gene, SEQ ID 5263.

XX

KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.

XX

OS Oryza sativa.

XX

PN WO2003000898-A1.

XX

PD 03-JAN-2003.

XX

PF 22-JUN-2001; 2001WO-IB001105.

XX

PR 22-JUN-2001; 2001WO-IB001105.

XX

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX

DR WPI; 2003-175290/17.

XX

PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.

XX

PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX

CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to

CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 2.8%; Score 44.6; DB 7; Length 2000;
Best Local Similarity 10.2%; Pred. No. 0.01;
Matches 95; Conservative 412; Mismatches 416; Indels 8; Gaps 4;

```
Qy      561 AGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAA 620
          :|:: :: :: : : : : : |: :: | : ::::|: : |:|: ::::|
Db      106 RGMRRSRMRWMGRYRRCARSGRMAGGSGRMMGGKSRMSYWMWCYARGCGSCKRKKSKGGS 165

Qy      621 TGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGAC 680
          :|: ::| :||: :: | :: : : : : || :: : :| ::|
Db      166 WGKTCRRGARGGSWSSGAKYKSGSMSKRMWMSSCGRSGCRRSAYSRYYGTSRKYGTYK 225

Qy      681 AGTCTGTAAC-AACAGTGAGTCGTTCTCTCCTCCTCCTGCGCAGGGCAGAGCCTGGAC 739
          |::: :| :|:: |: :: : :| :: : : : :|: : : :
Db      226 KMTYYASASRCMAYMTTTSYWACSSYTWCRRSKRRSMMWKMMRKMRSRSYGWYSWSYKMW 285

Qy      740 ATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTT 799
          : || : :: |: : | :: : :| : : ::::: : : :
Db      286 MCTAYKKSYYSRWCYMYRGGGWRGATRYWGRGYMSRMAMMYKKMYWYRGYKGMKRGWWAG 345

Qy      800 CCTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCT 859
          : : : : : : : :|: : : : : : : : : : : :
Db      346 RMMMRSMCRWSKACYYMRWRMWRMTRRRRWAKKSSRTSRRKKRKWCMRKRKYKMRGYSR 405

Qy      860 GGGGGCACAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAG 919
          : : : | : : : : : : :| : : : : : :|: : : : : :
Db      406 MRCKRARWMKRCRSGRAWKMGCRGCMTCRMKSYGMMRWKSWKRMASKYKWMRMYRWK 465

Qy      920 CAGTGTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAG 979
          :| :|:|: : : |: | : : |:|: | : : : : : : : :
Db      466 KKCSRTTMWGKTRGGMMGTMGRCRYKKRSGMKRKCRRRRWGRMYRMRWKRYYSARYTMR 525

Qy      980 GCATGCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCCAAGGACAACAGAGTGGTAC 1039
          || : : : : : : : :| : : : : : : : : : : : :
Db      526 YCARKKYSYSAARKARCWYRGKGYWAGMWMKRYKRMYYKMMWWYKRYKSKSWYCKMS 585

Qy      1040 ATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGG 1099
          :| :| |: ||: : : : : : : :|: | : : : :|:|:|:
Db      586 YYA--SCMKSARKAGAKMCKRSKMSAWSKSMRSSRKCRKCAKRSASAKRYAMGGMTSGS 643

Qy      1100 AGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTT 1159
          : : : :|: : : : : : :|: : : : : : : : : :
Db      644 RMSRWKSYTCYWRKWGSMKSTCTWYYMSKYTYAKYGSYWRYYR---RAWCMYMRWYYY 700

Qy      1160 AAAGCAACCGTGTGCGGCCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGC 1219
          : : : : : :|: : : || : : : :|: :| :|: : : : : : :|:
Db      701 RYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTWYKYC--KCSWKYRSMWYYSWVWAKTW 758

Qy      1220 AAATCCTCAAAGATGTCTCCTTGACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCA 1279
          : : : : : : : : : : : : : : : : : : : :
Db      759 KMWRRYATRMMWWMYRYSMKWYTWCTMWGYWYWWRTYMKMRMYWKCTKYWYWSATYW 818
```


QY 1280 GCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAGG 1339
 | | ::|:|:: | | : :: : : ::| : | : | : :
 Db 819 TGTWAAWWMMAKTKMRMGMTGAKTRGRARKARYWWKWATWCATKRWMTKGKGAKWAWTWMA 878

QY 1340 CTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTAAGTTGTAGAGAGGCAGCCA 1399
 : : :: : : : : ::|:|:| : |:: : :: : : | | : : |
 Db 879 KAWRKYYWSWMRAWYYYKTRRTRYKTCWWKARWGSWAYWRMWWKGSAKMWWMWKGGRWG 938

QY 1400 TGCATTTGGCATTGAATACAATCTGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACC 1459
 : :: : : | | | : : : | : : : : : | : | : :
 Db 939 WTKYWYWCCTTWKMACGRATKYMCCAGWWAMYSYSWTRTYWMRTWRWMWASSRTAKRMAR 998

QY 1460 AAAGTGAAATCTTGAGGAAAATCCCTGGAAA 1490
 ::::|:|: :: : | : | | | |
 Db 999 MMWKTRAWSKSYARAYWKMAGCACCTACACA 1029

RESULT 22

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

XX

AC ADA71938;

XX

DT 20-NOV-2003 (first entry)

XX

DE Rice gene, SEQ ID 5263.

XX

KW Plant; bacterial infection; fungal infection; viral infection; rice;
 KW gene; ds.

XX

OS Oryza sativa.

XX

PN WO2003000898-A1.

XX

PD 03-JAN-2003.

XX

PF 22-JUN-2001; 2001WO-IB001105.

XX

PR 22-JUN-2001; 2001WO-IB001105.

XX

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX

DR WPI; 2003-175290/17.

XX

PT Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.

XX

PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX

CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in

CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.

XX

SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 2.4%; Score 38.2; DB 7; Length 2000;
 Best Local Similarity 8.3%; Pred. No. 1.1;
 Matches 54; Conservative 292; Mismatches 300; Indels 1; Gaps 1;

Qy	233	AGATAAGGACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGA	292
		:::: : : :: : :::: : :: :::: : : : : : ::	
Db	651	RSMWYSKYSCSAKCCCKTRYMTSSYMSTGMYGMYSSYKSMWSKMSYMGKMTCTMYTSM	592
Qy	293	ACTGTTATCTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCT	352
		:: :: :::: : : : : : : : : : : : : : :::	
Db	591	KGSTRRSKMGRWSGMSRMYMRWWKKMRKRKMYRMWKCTWRRCMCYRWGYTMYTTSRSR	532
Qy	353	CCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACTTGCATTGCTT	412
		::: :: : : : : : : : : : : ::: : : : : : : : :	
Db	531	MMYTGRYKARYTSKRRYMWYKYRKYCWYYYGYMYMKCSYMMRYGYCKACKKCCYAMCWKA	472
Qy	413	CCTGCTAGCCATGGGTGAGCTGCCCTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACAT	472
		: : : : : : : : ::: : : : : : : : :	
Db	471	AYSGMMYWKYRKYSKWMRMSTKYMWSMWYKKCRSMKYGAKGCGYCKMWYTCYSGYMKWYT	412
Qy	473	CAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCA	532
		: : : : : : : : : : : : : : : : : : : : : :	
Db	411	YMGSYKYSRCYKYMRMYMYKGWYMMYYSAYSSMMTWYYYYAKYWKYWKRRGTMSWYGK	352
Qy	533	CAGCTTAGGTGTCTCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAA	592
		: : : : : : : : : : : : : : : : : : : : : :	
Db	351	SYKKKYCTWWCYMKCMRCYRWRKMMRKKTKYSKRCYCWRYATCYWCCCYRKRGWYSRRSM	292
Qy	593	GCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTG-GGTTGT	651
		: : : : : : : : : : : : : : : : : : : :	
Db	291	MRTAGKWKMRWSRWCRSYSWYKMYKKMWKKSYYMSYGWARSSGTWSRSAAKRTYKGYST	232
Qy	652	CTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTC	711
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	231	SRRAKMMRACRMYSACRRYSRTSYCGCSYCGSSKWKYMSKSCSMRMTCSSWCSCCYTCY	172
Qy	712	CTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCGCTTGCT	771
		: : : : : : : : : : : : : : : : : : : : : : :	
Db	171	YGAMCWSCCMSMMYMGSCGCYTRGWKWSKYSMCCCKKYCSCCTKYCSYTGYYRYCKWYKY	112
Qy	772	GCTTCTCACTGATTCTGCTCTCCCTTCTTGAAGTGGGCGCCACCTGTCTGTGTAGA	831
		:::: :: : : : : : : : : : : : : : : : : : : :	
Db	111	SYKCYCYCYWYMSYMRMYMKCMCSRSCSSWMSYAYCSTSSSRWMSMYAAKMGMCSS	52
Qy	832	TGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATG	878
		: : : :: : : : :: ::: : : : : : : : :	
Db	51	SGMYRMSKSKMYSKYSSCKYTGSKKCTKRKYCYWSSGYSMWCTS	5

RESULT 23

ADE57382/c

ID ADE57382 standard; DNA; 2692 BP.

XX

AC ADE57382;

XX

DT 29-JAN-2004 (first entry)

XX

DE Rat gene L38644, SEQ ID NO 3243.

XX

KW Rat; ds; gene; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

XX

OS Rattus norvegicus.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.

DR GENBANK; L38644.

XX

PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.

XX

PS Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 2692 BP; 738 A; 621 C; 711 G; 622 T; 0 U; 0 Other;

Query Match 2.4%; Score 37; DB 9; Length 2692;

Best Local Similarity 51.5%; Pred. No. 3.2;

Matches 85; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 703 TTCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTTAAACATGCCCTGCCTGAAG 762
| | ||| | | | | | ||| ||||| | | | | |
Db 235 TGCAACCTGATGTTTCCGGATTTGCCAGACTCTGGACGTTCCACCAGAACGTGGGAGATT 176

Qy 763 CCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTC 822
| | ||| ||| || ||||| | | | | | | | |
Db 175 CTCCCAGCCGCACCTCAAGGAATTCTGCGCCCTTCCAGCTCAGCCGATCGGAGACACGTC 116

Qy 823 CTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
| | ||||| | ||| | | | | | | | |
Db 115 TTCTCGAGATGGTTATAGCTCCATGCGGAGGTGAGGCGGCGGGAC 71

RESULT 24

AAH24065/c

ID AAH24065 standard; DNA; 4590 BP.

XX

AC AAH24065;

XX

DT 29-AUG-2001 (first entry)

XX

DE Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.

XX

KW Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
KW modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
KW lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
KW functional food; transgenic yeast; fat/lean ratio; food use; ds.

XX

OS *Saccharomyces cerevisiae*.

XX

FH Key Location/Qualifiers

FT misc_feature 10

FT /*tag= a

FT /note= "Represented as * in the specification"

FT misc_feature 3617

FT /*tag= b

FT /note= "Represented as * in the specification"

FT misc_feature 3649

```

FT          /*tag= c
FT          /note= "Represented as * in the specification"
FT  misc_feature  3679
FT          /*tag= d
FT          /note= "Represented as * in the specification"
FT  misc_feature  3819
FT          /*tag= e
FT          /note= "Represented as * in the specification"
FT  misc_feature  3862
FT          /*tag= f
FT          /note= "Represented as * in the specification"
FT  misc_feature  3864
FT          /*tag= g
FT          /note= "Represented as * in the specification"
FT  misc_feature  3888
FT          /*tag= h
FT          /note= "Represented as * in the specification"
FT  misc_feature  3890
FT          /*tag= i
FT          /note= "Represented as * in the specification"
FT  misc_feature  3912
FT          /*tag= j
FT          /note= "Represented as * in the specification"
FT  misc_feature  3914
FT          /*tag= k
FT          /note= "Represented as * in the specification"
FT  misc_feature  3938
FT          /*tag= l
FT          /note= "Represented as * in the specification"
FT  misc_feature  3939
FT          /*tag= m
FT          /note= "Represented as * in the specification"
FT  misc_feature  3941
FT          /*tag= o
FT          /note= "Represented as * in the specification"
FT  misc_feature  3943
FT          /*tag= p
FT          /note= "Represented as * in the specification"
FT  misc_feature  4361
FT          /*tag= q
FT          /note= "Represented as * in the specification"
XX
PN  WO200133977-A1.
XX
PD  17-MAY-2001.
XX
PF  06-NOV-2000; 2000WO-AU001362.
XX
PR  05-NOV-1999; 99AU-00003875.
XX
PA  (META-) METABOLIC PHARM LTD.
XX
PI  Belyea CI, Ng FM, Vaughan P;
XX
DR  WPI; 2001-328876/34.
XX
PT  New organisms containing nucleic acid encoding a growth hormone fragment

```

PT which modulates lipid metabolism are useful to produce dietary aids for
PT obesity and in the meat production industry.

XX

PS Disclosure; Page 48-50; 54pp; English.

XX

CC The invention relates to novel transgenic organisms useful in the
CC production of functional food and drink products for the treatment or
CC prevention of obesity via the regulation of lipid metabolism. The
CC organisms comprise a polynucleotide encoding a growth hormone fragment
CC capable of stimulating the activity of hormone-sensitive lipase (the key
CC enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
CC enzyme in lipogenesis). The growth hormone fragment preferably contains
CC at least the disulphide-bonded loop of a mammalian growth hormone (but is
CC not the full-length growth hormone) and is optionally linked to an
CC epitope tag or heterologous fusion protein partner. The transgenic
CC organism may be a microorganism used to produce a fermented product
CC (e.g., yeast), or an edible plant or animal or cell thereof. Food or
CC drink made using methods of the invention are used to modify fat/lean
CC ratio, lipid metabolism or food use in a mammal. In particular, the food
CC or drink products may be used to treat or prevent obesity, particularly
CC in humans, and may also be used to improve the fat/lean ration of
CC livestock raised for meat production. In the exemplification of the
CC invention, the human growth hormone (hGH) fragment analogue AOD9604 was
CC expressed in yeast, optionally fused to the FLAG epitope (AAB73625). The
CC present sequence is described as a DNA sequence from yeast in the
CC sequence listing, but is not further referred to in the specification

XX

SQ Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 0 U; 2896 Other;

Query Match 2.4%; Score 37; DB 5; Length 4590;

Best Local Similarity 9.3%; Pred. No. 4.3;

Matches 60; Conservative 298; Mismatches 285; Indels 0; Gaps 0;

Qy 210 ACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATAAT 269
:: : :: : :: : | : : | : : | : : : :: | | : : : : | : :
Db 4375 VYSYYTDSYRYANAYHHHVNTCHAADGMGTDDAYCHSYHYWASYGKHSRHNWGSNHN 4316
Qy 270 GGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCAGGGCTGGGTAGGAT 329
:: : : : : : : : : | | : : : : : : : : : | : : :
Db 4315 SRHNWSSDDSRHNWSRHNWAHGSSATKASGHHYWHASSVKDHSVDDWNYGYTVKRSN 4256
Qy 330 CGGACAGGGCACTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACT 389
: : : : : : : : : : : : : : | : : : : : : : : |
Db 4255 TKYWNSKACKSSWMSWMSMYHSTBTSRYBGYATKAGSRHNWHSTBTSRYBGYATKAGS 4196
Qy 390 AGAAAATTCACATTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCA 449
: : : | : : : | : : | : : : : : : : : : : : : : : | : :
Db 4195 RHNWHSTBTSRYBGYATKAGSRHNWHSTBTSRYBGYATKAGSRHNWHSTBTSRYBGYATK 4136
Qy 450 GAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCG 509
: : : : : : : : | : : : | : : : : : : : : : : : : : : | : : :
Db 4135 AGSRHNWHSTBTSRYBGYATKAGSRHNWGHMSRHNWKDSVKSRHNWNMYHVVCARRYWBH 4076
Qy 510 GTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGG 569
: : : : : : : : | : : : : : : : : | : : : : : : : : : : : : : :
Db 4075 VHNMRWMKKKMGKKHGSYVKNNYVKNCTYYAYYHTDANDTYCTYTATHTDMGCNHTDDD 4016

PT Genes and their expression products cloned from human cDNA libraries for
PT treatment and diagnosis of diseases associated with their expression.
XX
PS Claim 1(a); Page 123-127; 238pp; Japanese.
XX
CC The invention relates to DNA encoding polypeptides directly cloned from
CC cDNA libraries originating in adult whole brain, human tonsil, human
CC adult hippocampus and human foetal whole brain. Polypeptides and
CC polynucleotides of the invention may be used in the investigation of
CC differential expression of the DNA sequences in normal subjects and
CC disease patients. They may also be used in the production of antibodies,
CC oligonucleotide probes and DNA chips for diagnosis and identification of
CC drugs for treatment of diseases with which the DNA sequences are
CC associated. The sequences given in records ABN83954-ABN83984 represent
CC human gene sequences of the invention
XX
SQ Sequence 6843 BP; 1812 A; 1693 C; 1483 G; 1855 T; 0 U; 0 Other;

Query Match 2.3%; Score 36.8; DB 6; Length 6843;
Best Local Similarity 49.0%; Pred. No. 6.3;
Matches 98; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 741 TTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTC 800
|||| | | | || || | | | | | | | | | | | |
Db 689 TTAATATACTCTATGGATGACCCAGCAAGTTTGCTGTTTCAGAATCCTCCTCTTCTGTTT 748
Qy 801 CTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTG 860
|||| | | || | || | || | || | || | | | |
Db 749 TTTGAACTTTCGAAAACAAAAGATATGCTGGGAGACGCGGCCCTAGAGTGTGCTTACTC 808
Qy 861 GGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGC 920
|| | || | || || | || | || | || | | | |
Db 809 CAGGTCCTTGATTGTCCAGACTGTGGAGGGGAAGGGCAGATCTATGCCAAGAGGGGAAC 868
Qy 921 AGTGTGCAGGTGTACCATCT 940
|| || | | || ||
Db 869 AGGCTGTAGAGGCCACAGCT 888

RESULT 26

AAS93276/c

ID AAS93276 standard; cDNA; 541 BP.

XX

AC AAS93276;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #29080.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG29089.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 29080; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 541 BP; 121 A; 154 C; 168 G; 98 T; 0 U; 0 Other;

Query Match 2.3%; Score 36.2; DB 5; Length 541;
Best Local Similarity 50.3%; Pred. No. 2.2;
Matches 83; Conservative 2; Mismatches 80; Indels 0; Gaps 0;

Qy 1174 GGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCCAAATCCTCAAAGAT 1233
| ||||| |||| | | | | | | | | | | | | | | | |
Db 223 GTGCCTTGCTGGATCCTCTGCGCCGCGCAGATGCCGTAGGCCAGGCCGGGCACAGTACTG 164

Qy 1234 GTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGC 1293
|| | ||| | | | | | | | | | | | | | | | |
Db 163 GTGCAGAGGCACACCTCGCGAGGCAGGTCCCGCAGCCACTCCGGCAGCTCCGGCGGGGGC 104

Qy 1294 CTGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAG 1338

Db 103 GCGGCGGCCGCGAGCGCTGCTGGTGCCCAAGCCGGCGCAGCGAG 59

RESULT 27

ADA08012/c

ID ADA08012 standard; cDNA; 2236 BP.

XX

AC ADA08012;

XX

DT 06-NOV-2003 (first entry)

XX

DE cDNA encoding human PR Family member 3b (PFM3b).

XX

KW Human; PR Family Member 1; PR Family Member 2; PR Family Member 3;
KW PR Family Member 4; PR Family Member 5; PFM1; PFM2; PFM3; PFM4; PFM5;

KW PFM PR domain; PFM zinc finger domain; PFM ZF domain;

KW modulation of cell growth; cancer; cell degeneration disease;

KW Alzheimer's disease; Parkinson's disease;

KW insulin-dependent diabetes mellitus; IDDM; neuroprotective;

KW antiparkinsonian; antidiabetic; cytostatic; gene; ss.

XX

OS Homo sapiens.

XX

PN US6586579-B1.

XX

PD 01-JUL-2003.

XX

PF 03-SEP-1999; 99US-00389956.

XX

PR 03-SEP-1999; 99US-00389956.

XX

PA (BURN-) BURNHAM INST.

XX

PI Huang S;

XX

DR WPI; 2003-669568/63.

DR P-PSDB; ADA08013.

XX

PT New PR Family Member 2 oligonucleotide, useful for preparing a
PT composition for modulating cell growth for treating cancer or diseases of
PT cell degeneration, e.g., Alzheimer's disease or insulin-dependent
PT diabetes mellitus.

XX

PS Example 3; Fig 6A; 95pp; English.

XX

CC The present invention relates to the isolation of human and mouse PR
CC Family Member (PFM) proteins, and the polynucleotide sequences encoding
CC them. Also disclosed are PFM PR and PFM zinc finger (ZF) domains, and the
CC polynucleotide sequences encoding them. The invention also discloses PFM
CC oligonucleotides and methods for detecting a PFM polynucleotide sequence
CC in a sample. The PFM polypeptide and polynucleotide sequences are useful
CC for preparing a composition for modulating cell growth for treating
CC cancer or diseases of cell degeneration, e.g. as Alzheimer's disease,
CC Parkinson's disease or insulin-dependent diabetes mellitus (IDDM). The
CC present sequence represents a PFM polynucleotide sequence of the
CC invention.

XX

SQ Sequence 2236 BP; 516 A; 656 C; 580 G; 484 T; 0 U; 0 Other;

Query Match 2.3%; Score 36.2; DB 8; Length 2236;
Best Local Similarity 50.3%; Pred. No. 5.1;
Matches 83; Conservative 2; Mismatches 80; Indels 0; Gaps 0;

```
Qy      1174 GGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGAT 1233
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      713 GTGCCTTGCTGGATCCTCTGCGCCGCGCAGATGCCGTAGGCCAGGCCGGGCACAGTACTG 654

Qy      1234 GTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGC 1293
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      653 GTGCAGAGGCACACCTCGCGAGGCAGGTCCCGCAGCCACTCCGGCAGCTCCGGCGGGGGC 594

Qy      1294 CTGGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAG 1338
          | | | | : : | | | | | | | | | | | | | | | | | |
Db      593 GCGGCGGCCGCGAGCGCTGCTGGTGCCACAAGCCGGCGCAGCGAG 549
```

RESULT 28

ADA08010/c

ID ADA08010 standard; cDNA; 2488 BP.

XX

AC ADA08010;

XX

DT 06-NOV-2003 (first entry)

XX

DE cDNA encoding human PR Family member 3a (PFM3a).

XX

KW Human; PR Family Member 1; PR Family Member 2; PR Family Member 3;

KW PR Family Member 4; PR Family Member 5; PFM1; PFM2; PFM3; PFM4; PFM5;

KW PFM PR domain; PFM zinc finger domain; PFM ZF domain;

KW modulation of cell growth; cancer; cell degeneration disease;

KW Alzheimer's disease; Parkinson's disease;

KW insulin-dependent diabetes mellitus; IDDM; neuroprotective;

KW antiparkinsonian; antidiabetic; cytostatic; gene; ss.

XX

OS Homo sapiens.

XX

PN US6586579-B1.

XX

PD 01-JUL-2003.

XX

PF 03-SEP-1999; 99US-00389956.

XX

PR 03-SEP-1999; 99US-00389956.

XX

PA (BURN-) BURNHAM INST.

XX

PI Huang S;

XX

DR WPI; 2003-669568/63.

DR P-PSDB; ADA08011.

XX

PT New PR Family Member 2 oligonucleotide, useful for preparing a

PT composition for modulating cell growth for treating cancer or diseases of

AA119738

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR

PR

PR

PR

PR

PR

XX

XX

XX

XX

XX

XX

XX

— 2 —

Best Local Similarity 62.5%; Pred. No. 2.5;

Plasmones de superficie, constituidos por los plasmones de volumen de los metales, se propagan a lo largo de la interfase entre un metal y un dieléctrico. Los plasmones de superficie se propagan a lo largo de la interfase entre un metal y un dieléctrico. Los plasmones de superficie se propagan a lo largo de la interfase entre un metal y un dieléctrico.

Db 206 CTTAATACCTCCAGCAACCAGTTGTGACAATACATGCAAAGAGTGCAAAGTCTTGTCCAC 265

Qy 1511 CACGTAAGTGCCTTCTTTGCTTGTGTTGA 1538
 ||| | || | || || | || |||||

Db 266 GACGGATGTNCTTTTTTTTTTTTTTTTGA 293

RESULT 31

AAH13437/c

ID AAH13437 standard; cDNA; 493 BP.

XX

AC AAH13437;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA clone (3'-primer) SEQ ID NO:10272.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX

PS Claim 3; SEQ ID NO 10272; 2537pp + Sequence Listing; English.

XX

CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

XX
 SQ Sequence 493 BP; 112 A; 142 C; 138 G; 79 T; 0 U; 22 Other;

Query Match 2.3%; Score 35.8; DB 4; Length 493;
 Best Local Similarity 57.5%; Pred. No. 2.8;
 Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 780 CTGATTCTGCTCTCCCCTTCCTTGACTCGCCACCACCTGTCCTGTGTAGATGGAGAAG 839
 || | | | | | | | | | | | | | | | | | | | |
 Db 267 CTNAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCTCCCTNTGGGAATTGGGCAG 208

 Qy 840 GCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
 | | | | | | | | | | | | | | | | | | | |
 Db 207 CCCTGGGCAGTTGTACTCATGGGGGCTTAANATGCAGCTACCTCAG 162

RESULT 32

ABN73379/c

ID ABN73379 standard; cDNA; 639 BP.

XX

AC ABN73379;

XX

DT 03-JUL-2002 (first entry)

XX

DE Bovine embryonic germ (EG) cell cDNA EST 000203a CONTIG 63.

XX

KW Bovine; Bos taurus; EST; expressed sequence tag; totipotence;

KW development; gene; ss.

XX

OS Bos taurus.

XX

PN WO200194550-A2.

XX

PD 13-DEC-2001.

XX

PF 07-JUN-2001; 2001WO-US018576.

XX

PR 07-JUN-2000; 2000US-0209874P.

PR 06-JUN-2001; 2001US-00876143.

XX

PA (INFI-) INFIGEN INC.

XX

PI Eilertsen KJ, Pfister-Genskow M, Childs L;

XX

DR WPI; 2002-351289/38.

XX
PT An expressed sequence tag (EST), the expression of which, or its
PT complementary sequence, in a cell identifies the cell as a
PT developmentally competent or incompetent cell.
XX
PS Example 16; Page 161; 584pp; English.
XX
CC The present invention describes an expressed sequence tag (EST), where
CC the EST is an isolated, enriched, or purified nucleic acid sequence
CC representing all or part of a gene, the expression of which, or its
CC complementary sequence, in a cell identifies the cell as a
CC developmentally competent or incompetent cell. Molecules which induce
CC developmental competence in a cell line are useful for inducing
CC totipotency in one or more cells. Molecules which induce developmental
CC incompetence in a cell line are useful for preventing a full term
CC pregnancy in an animal and inhibiting totipotency. The molecules are also
CC useful for treating a disease in an animal by inducing development of one
CC or more cells of the animal into a specific cell type. The present
CC sequence represents a bovine EST which is given in the exemplification of
CC the present invention
XX
SQ Sequence 639 BP; 149 A; 184 C; 142 G; 156 T; 0 U; 8 Other;

Query Match 2.3%; Score 35.8; DB 6; Length 639;
Best Local Similarity 51.5%; Pred. No. 3.3;
Matches 69; Conservative 4; Mismatches 61; Indels 0; Gaps 0;

QY 1185 GAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTA 1244
|||: |: || ||| || ||||| ||| || ||| |
Db 425 GAAVCTVCCGTCCTGCAAGTCAGAGTGGGACACACAAAGTCTGCTGTTTGTCTGGCAGANC 366

QY 1245 CATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCS 1304
|| ||| || || || || || || || || || || |||||: :
Db 365 CACTTCAAGTACGNAATNAAGAGCAGCATGAAGAGATCTGGTGAAATTCTGGGGGGGAG 306

QY 1305 GGGGCTCCTGTACT 1318
|| ||| ||
Db 305 AAGGGAGCTGCTCT 292

RESULT 33

ABL79659

ID ABL79659 standard; cDNA; 257 BP.

XX

AC ABL79659;

XX

DT 17-MAY-2002 (first entry)

XX

DE Human ovarian cancer related cDNA clone SEQ ID NO:2637.

XX

KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200192581-A2.

XX

PD 06-DEC-2001.

XX
 PF 29-MAY-2001; 2001WO-US017756.
 XX
 PR 26-MAY-2000; 2000US-0207484P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Algate PA, Harlocker SL, Jones R;
 XX
 DR WPI; 2002-122075/16.
 XX
 PT Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.
 XX
 PS Claim 1; SEQ ID NO 2637; 489pp; English.
 XX
 CC The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
 CC or antigen presenting cells that express (II). (I) has cytostatic
 CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 CC detecting ovarian cancer in a patient's biological sample preferably
 CC serum or ovarian tissue. The method comprises contacting a biological
 CC sample from a patient with (IV), detecting the amount of polynucleotide
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff
 CC value and thereby detecting ovarian cancer in the patient, where the
 CC amount of polynucleotide hybridising to (IV) is detected preferably by
 CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
 CC useful for stimulating and/or expanding T cells specific for an ovarian
 CC tumour protein comprising contacting T cells with (III) or (II). (III) is
 CC useful in design and preparation of ribozyme molecules for inhibiting
 CC expression of the tumour polypeptides and proteins in tumour cells; and
 CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
 CC library using well known techniques
 XX
 SQ Sequence 257 BP; 28 A; 81 C; 83 G; 65 T; 0 U; 0 Other;

Query Match 2.3%; Score 35.6; DB 6; Length 257;
 Best Local Similarity 58.5%; Pred. No. 2.2;
 Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 780 CTGATTTCTGCTCTCCCTTCCTTGACTCGCCACCACCTGTCCTGTGTAGATGGAGAAG 839
 ||| | | |||| | | ||| | | ||| ||| ||| ||| |||
 Db 12 CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCCTCTGGGAATTGGGCAG 71
 Qy 840 GCTCGGAGAGTGGGGGTGCTGGGGGCACAAATGGAATGAACACTG 885
 | || ||| | | ||||| || ||| | ||| |
 Db 72 CCCTGGGCAGTTGTACTCATGGGGCTTAAGATGCAGCTACCTCAG 117

RESULT 34
 AAQ60875/c
 ID AAQ60875 standard; DNA; 313 BP.

XX
 AC AAQ60875;
 XX
 DT 25-MAR-2003 (revised)
 DT 16-MAR-1994 (first entry)
 XX
 DE Human brain Expressed Sequence Tag EST00969.
 XX
 KW Gene transcription product; genetic markers; tagging; in vivo;
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9316178-A2.
 XX
 PD 19-AUG-1993.
 XX
 PF 12-FEB-1993; 93WO-US001294.
 XX
 PR 12-FEB-1992; 92US-00837195.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX
 PI Venter CJ, Adams MD, Moreno RF;
 XX
 DR WPI; 1993-272882/34.
 XX
 PT Enriched oligonucleotides and corresp. sequences - used as markers for
 PT human genes transcribed in-vivo, facilitate tagging of most human genes.
 XX
 PS Example 4; Page 404; 500pp; English.
 XX
 CC The Expressed Sequence Tag was isolated from a human brain cDNA library
 CC as part of a large set of ESTs which can be used as markers for human
 CC genes transcribed in vivo. They can be used to facilitate tagging of most
 CC human genes, for mapping locations of expressed genes on chromosomes, for
 CC individual or forensic identification, for mapping locations of disease-
 CC associated genes, for identification of tissue type, and for prepn. of
 CC antisense sequences, probes and constructs. EST00969 has a "poor" coding
 CC probability as evaluated using the coding-region prediction program CRM.
 CC See also AAQ59041-Q61440. (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 313 BP; 75 A; 106 C; 90 G; 41 T; 0 U; 1 Other;

Query Match 2.3%; Score 35.6; DB 2; Length 313;
 Best Local Similarity 58.5%; Pred. No. 2.5;
 Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 780 CTGATTTCTGCTCTCCCCTTCCTTGACTCGCCACACCTGTCCTGTGTAGATGGAGAAG 839
 ||| | | ||||| | | || | | | | | | | | | |
 Db 250 CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCCCTCTGGGAATTGGGCAG 191
 Qy 840 GCTCGGAGAGTGGGGGTGCTGGGGGCACAAATGGAATGAACACTG 885
 | || ||| | | ||||| || || | | | | |
 Db 190 CCCTGGGCAGTTGTACTCATGGGGGCTTAAGATGCAGCTACCTCAG 145

RESULT 35

ABN94395/c

ID ABN94395 standard; DNA; 330 BP.

XX

AC ABN94395;

XX

DT 13-AUG-2002 (first entry)

XX

DE Gene #893 used to diagnose liver cancer.

XX

KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumour; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX

OS Homo sapiens.

XX

PN WO200229103-A2.

XX

PD 11-APR-2002.

XX

PF 02-OCT-2001; 2001WO-US030589.

XX

PR 02-OCT-2000; 2000US-0237054P.

XX

PA (GENE-) GENE LOGIC INC.

XX

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX

DR WPI; 2002-426119/45.

XX

PT Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.

XX

PS Claim 1; SEQ ID NO 893; 298pp; English.

XX

CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumour in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 330 BP; 77 A; 104 C; 88 G; 61 T; 0 U; 0 Other;

Query Match 2.3%; Score 35.6; DB 6; Length 330;

Best Local Similarity 58.5%; Pred. No. 2.6;

Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 780 CTGATTCTGCTCTCCCCTCCTTGACTCGCCACACCTGTCCTGTGTAGATGGAGAAG 839
 |||| | | ||||| | | || | | || | | || | | || | | ||
 Db 282 CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCACCTCTGGGAATTGGGCAG 223
 Qy 840 GCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
 | || ||| | | ||||| || ||| | | || |
 Db 222 CCCTGGGCAGTTGTACTCATGGGGGCTTAAGATGCAGCTACCTCAG 177

RESULT 36

ABL87579/c

ID ABL87579 standard; cDNA; 440 BP.

XX

AC ABL87579;

XX

DT 17-MAY-2002 (first entry)

XX

DE Human ovarian cancer related cDNA clone SEQ ID NO:10557.

XX

KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200192581-A2.

XX

PD 06-DEC-2001.

XX

PF 29-MAY-2001; 2001WO-US017756.

XX

PR 26-MAY-2000; 2000US-0207484P.

XX

PA (CORI-) CORIXA CORP.

XX

PI Algate PA, Harlocker SL, Jones R;

XX

DR WPI; 2002-122075/16.

XX

PT Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.

XX

PS Claim 1; SEQ ID NO 10557; 489pp; English.

XX

CC The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
 CC or antigen presenting cells that express (II). (I) has cytostatic
 CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 CC detecting ovarian cancer in a patient's biological sample preferably
 CC serum or ovarian tissue. The method comprises contacting a biological
 CC sample from a patient with (IV), detecting the amount of polynucleotide
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff
 CC value and thereby detecting ovarian cancer in the patient, where the

CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques

XX

SQ Sequence 440 BP; 97 A; 154 C; 131 G; 56 T; 0 U; 2 Other;

Query Match 2.3%; Score 35.6; DB 6; Length 440;
Best Local Similarity 58.5%; Pred. No. 3.1;
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 780 CTGATTTCTGCTCTCCCTTCCTTGACTCGCCACACCTGTCCTGTGTAGATGGAGAAG 839
|||| | | ||||| | | ||| | | ||| ||| || | | ||
Db 261 CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCTCCCCTCTGGGAATTGGGCAG 202

Qy 840 GCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
| || ||| | | ||||| || ||| | | ||| |
Db 201 CCCTGGGCAGTTGTACTCATGGGGCTTAAGATGCAGCTACCTCAG 156

RESULT 37

AAC76886/c

ID AAC76886 standard; cDNA; 1333 BP.

XX

AC AAC76886;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF2441 polynucleotide sequence SEQ ID NO:4881.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US008621.

XX

PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach M;

XX

DR WPI; 2000-602362/57.

DR P-PSDB; AAB42677.

XX

PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.

XX

PS Claim 5; Page 4060-4061; 5507pp; English.

XX

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX

SQ Sequence 1333 BP; 268 A; 380 C; 396 G; 286 T; 0 U; 3 Other;

Query Match 2.3%; Score 35.6; DB 3; Length 1333;

Best Local Similarity 46.7%; Pred. No. 5.8;

Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```
Qy      646 GGTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTTC 705
          || | ||| | | | | | | | | | | | | | | | |
Db      1285 GGCAGGCTGTTCTCTGGTTCCAACACTTGCCACAGGATCTCTAAAGACCCAGGAATGG 1226

Qy      706 CTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCG 765
          || | | | | | | | | | | | | | | | | | |
Db      1225 GGGCTATTGCCAGGGGTTAGAAGAGAACCAGGTCCCAAGGGCATGGTGGGCGGGCAGATG 1166

Qy      766 CTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTTGACTCGCCACACCTGTCCTG 825
          || | | | | | | | | | | | | | | | | | |
Db      1165 GTTCCAGAGCCTTAGAGATTCATAGGTTCTTCCTCCTCCACCAGCTGCTCCGAGGGCCTG 1106
```

QY 826 TGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
 || || ||| |||| | ||| ||| | ||| || || | | | |
 Db 1105 TGGGGAGGGACAAGGGTGGGATGCTGGAGCACCAGGGCTGCAGCAAGGGCCTTAGCTAAG 1046
 QY 886 CT 887
 ||
 Db 1045 CT 1044

RESULT 38

AAH18291

ID AAH18291 standard; cDNA; 2474 BP.

XX

AC AAH18291;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA sequence SEQ ID NO:18274.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX

PS Claim 8; SEQ ID NO 18274; 2537pp + Sequence Listing; English.

XX

CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

XX

SQ Sequence 2474 BP; 468 A; 861 C; 716 G; 429 T; 0 U; 0 Other;

Query Match 2.3%; Score 35.6; DB 4; Length 2474;
 Best Local Similarity 58.5%; Pred. No. 8.4;
 Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 780 CTGATTTCTGCTCTCCCTTCCTTGACTCGCCACCACCTGTCCTGTGTAGATGGAGAAG 839
 |||| | | ||||| | | || | | | | | | | | | |
 Db 2208 CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCTCCCTCTGGGAATTGGGCAG 2267
 QY 840 GCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
 | || ||| | | ||||| || || | | | |
 Db 2268 CCCTGGGCAGTTGTACTCATGGGGGCTTAAGATGCAGCTACCTCAG 2313

RESULT 39

AAS42019

ID AAS42019 standard; DNA; 21632 BP.

XX

AC AAS42019;

XX

DT 17-DEC-2001 (first entry)

XX

DE Genomic sequence #335 encoding novel human enzyme polypeptide.

XX

KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

KW ligase; hyperproliferative disorder; immunodeficiency disorder;

KW autoimmune disorder; neurological disorder; metabolic disorder;

KW inflammatory disorder; cardiovascular disorder; reproductive disorder;

KW blood-related disorder; infectious disorder; gene therapy; cytostatic;

KW anti arthritic; nephrotropic; anticoagulant; ds.

XX

OS Homo sapiens.

XX

PN WO200155301-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US001239.

XX

PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.

PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-465566/50.

XX

PT Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT treating neural, immune system, muscular, reproductive, pulmonary,
PT cardiovascular, renal, proliferative disorders and cancerous diseases.

XX

PS Disclosure; SEQ ID NO 2145; 1180pp; English.

XX

CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC infectious disorders (e.g. Influenza). The polynucleotides of the
CC invention can also be used in gene therapy. AAS41685-AAS42192 represent
CC DNA sequences encoding for the novel human enzyme polypeptides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 21632 BP; 4707 A; 6164 C; 6039 G; 4722 T; 0 U; 0 Other;

Query Match 2.3%; Score 35.6; DB 4; Length 21632;
Best Local Similarity 46.7%; Pred. No. 30;
Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 646 GGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAAACAACAGTGAGTCGTTC 705
 || | ||| | | | | | | | | | | | |
 Db 10501 GGCAGGCTGTTCTCTGGTTCCAAC TACTTGCCACAGGATCTCTAAAGACCCAGGAATGG 10560
 Qy 706 CTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCG 765
 || | | | | | | | | | | | | | | | |
 Db 10561 GGGCTATTGCCAGGGGTTAGAAGAGAACCAGGTCCCAAGGGCATGGTGGGCGGGCAGATG 10620
 Qy 766 CTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCTG 825
 || | | | | | | | | | | | | | | | |
 Db 10621 GTTCCAGAGCCTTAGAGATT CATAGGTTCTTCCTCCTCCACCAGCTGCTCCGAGGGCCTG 10680
 Qy 826 TG TAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
 || || || | || | || | || | || | || | || |
 Db 10681 TGGGGAGGGACAAGGGTGGGATGCTGGAGCACCAGGGCTGCAGCAAGGGCCTTAGCTAAG 10740
 Qy 886 CT 887
 ||
 Db 10741 CT 10742

RESULT 40

AAC89560/c

ID AAC89560 standard; DNA; 122186 BP.

XX

AC AAC89560;

XX

DT 08-MAR-2001 (first entry)

XX

DE Human histone deacetylase HDAC-D coding sequence.

XX

KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;

KW HDAC-D; cell cycle; tumorigenesis; cancer; inhibitor; antisense;

KW gene therapy; ds.

XX

OS Homo sapiens.

XX

PN WO200071703-A2.

XX

PD 30-NOV-2000.

XX

PF 03-MAY-2000; 2000WO-IB001252.

XX

PR 03-MAY-1999; 99US-0132287P.

XX

PA (METH-) METHYLGENE INC.

XX

PI Macleod AR, Li Z, Besterman JM;

XX

DR WPI; 2001-016407/02.

XX

PT Antisense oligonucleotide that inhibits expression of a histone
 PT deacetylase, useful for treating and/or alleviating the symptoms of
 PT neoplasia, or for inhibiting neoplastic cell growth in an animal.

XX

PS Disclosure; Page 89-125; 125pp; English.

XX

CC The present invention provides inhibitors of histone deacetylase enzymes
CC such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These
CC inhibitors may be antisense strands or they may be compounds identified
CC by contacting the enzyme with the compound and measuring the resulting
CC enzyme activity. These inhibitors are useful for treating cancers and for
CC identifying which histone deacetylase is involved in a neoplasia

XX

SQ Sequence 122186 BP; 29016 A; 31077 C; 32425 G; 29668 T; 0 U; 0 Other;

Query Match 2.3%; Score 35.6; DB 4; Length 122186;
Best Local Similarity 46.7%; Pred. No. 81;
Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 646 GGTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTC 705
|| | ||| | | | | | | | | | | | | | |
Db 107424 GGCAGGCTGTTCTCTGGTTCCAACACTTGTCCACAGGATCTCTAAAGACCCAGGAATGG
107365

Qy 706 CTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCG 765
|| | | | | | | | | | | | | | | | | |
Db 107364 GGGCTATTGCCAGGGGTTAGAAGAGAACCAGGTCCCAAGGGCATGGTGGGCGGGCAGATG
107305

Qy 766 CTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTTGACTCGCCCACCACCTGTCCTG 825
|| | | | | | | | | | | | | | | | | |
Db 107304 GTTCCAGAGCCTTAGAGATTATAGGTTCTTCCTCCTCCACCAGCTGCTCCGAGGGCCTG
107245

Qy 826 TGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
|| || ||| |||| | ||| ||| | | ||| || || | | | |
Db 107244 TGGGGAGGGACAAGGGTGGGATGCTGGAGCACCAGGGCTGCAGCAAGGGCCTTAGCTAAG
107185

Qy 886 CT 887
||
Db 107184 CT 107183

RESULT 41

ABN73289/c

ID ABN73289 standard; cDNA; 639 BP.

XX

AC ABN73289;

XX

DT 03-JUL-2002 (first entry)

XX

DE Bovine embryonic germ (EG) cell cDNA EST 000203a CONTIG 63.

XX

KW Bovine; Bos taurus; EST; expressed sequence tag; totipotence;
KW development; gene; ss.

XX

OS Bos taurus.

XX

PN WO200194550-A2.

XX

PD 13-DEC-2001.

XX
 PF 07-JUN-2001; 2001WO-US018576.
 XX
 PR 07-JUN-2000; 2000US-0209874P.
 PR 06-JUN-2001; 2001US-00876143.
 XX
 PA (INFI-) INFIGEN INC.
 XX
 PI Eilertsen KJ, Pfister-Genskow M, Childs L;
 XX
 DR WPI; 2002-351289/38.
 XX
 PT An expressed sequence tag (EST), the expression of which, or its
 PT complementary sequence, in a cell identifies the cell as a
 PT developmentally competent or incompetent cell.
 XX
 PS Example 16; Page 143-144; 584pp; English.
 XX
 CC The present invention describes an expressed sequence tag (EST), where
 CC the EST is an isolated, enriched, or purified nucleic acid sequence
 CC representing all or part of a gene, the expression of which, or its
 CC complementary sequence, in a cell identifies the cell as a
 CC developmentally competent or incompetent cell. Molecules which induce
 CC developmental competence in a cell line are useful for inducing
 CC totipotency in one or more cells. Molecules which induce developmental
 CC incompetence in a cell line are useful for preventing a full term
 CC pregnancy in an animal and inhibiting totipotency. The molecules are also
 CC useful for treating a disease in an animal by inducing development of one
 CC or more cells of the animal into a specific cell type. The present
 CC sequence represents a bovine EST which is given in the exemplification of
 CC the present invention
 XX
 SQ Sequence 639 BP; 149 A; 184 C; 142 G; 156 T; 0 U; 8 Other;

Query Match 2.3%; Score 35.4; DB 6; Length 639;
 Best Local Similarity 51.5%; Pred. No. 4.4;
 Matches 69; Conservative 2; Mismatches 63; Indels 0; Gaps 0;

Qy 1185 GAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTA 1244
 ||| | || ||| || ||||| ||| || ||| |
 Db 425 GAANCTNCCGTCCTGCAAGTCAGAGTGGGACACACAAAGTCTGCTGTTTGTCTGGCAGANC 366
 Qy 1245 CATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCS 1304
 || ||| | || | ||| | || ||| || ||||| : :
 Db 365 CACTTCAAGTACGNAATNAAGAGCAGCATGAAGAGATCTGGTGAAATCTGGGGGGGAG 306
 Qy 1305 GGGGCTCCTGTACT 1318
 || ||| ||
 Db 305 AAGGGAGCTGCTCT 292

RESULT 42

AAL18444

ID AAL18444 standard; cDNA; 414 BP.

XX

AC AAL18444;

XX

AAL10158

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 14-MAR-2000; 2000US-0189167P.

PR 29-MAR-2000; 2000US-0193480P.

PR 09-JUN-2000; 2000US-0211315P.

PR 25-JUL-2000; 2000US-0220534P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX

DR WPI; 2001-451856/48.

XX

PT New peptide useful as a marker for the diagnosis of breast cancer.

XX

PS Claim 1; Page 495; 3695pp; English.

XX

CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity

XX

Sequence 416 BP; 137 A; 92 C; 94 G; 93 T; 0 U; 0 Other;

Query Match 2.2%; Score 35.2; DB 4; Length 416;

Best Local Similarity 62.5%; Pred. No. 4;

Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qv 1451 CCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAGTCCTGCCTAA 1510

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Db 221 CTTAATACCTCCAGCAACCAGTTGTGACAATACATGCAAAGAGTGCAAAGTCTTGTCCAC 280

Qy 1511 CACGTAAGTGCCTTCTTTGCTTGTGTTGA 1538
 ||| | || | || || | || ||||
 Db 281 GACGGATGTTCTTTTTTTTTTTTTTTGA 308

RESULT 44

AAL09772

ID AAL09772 standard; cDNA; 459 BP.

XX

AC AAL09772;

XX

DT 07-DEC-2001 (first entry)

XX

DE Human breast cancer expressed polynucleotide 2229.

XX

KW Human; breast cancer; cell marker; cytostatic; ss.

XX

OS Homo sapiens.

XX

PN WO200151628-A2.

XX

PD 19-JUL-2001.

XX

PF 10-JAN-2001; 2001WO-US000798.

XX

PR 14-JAN-2000; 2000US-0176077P.

PR 14-MAR-2000; 2000US-0189167P.

PR 24-MAR-2000; 2000US-0192099P.

PR 29-MAR-2000; 2000US-0193480P.

PR 15-MAY-2000; 2000US-0205230P.

PR 09-JUN-2000; 2000US-0211315P.

PR 25-JUL-2000; 2000US-0220534P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX

DR WPI; 2001-451856/48.

XX

PT New peptide useful as a marker for the diagnosis of breast cancer.

XX

PS Claim 1; Page 428-429; 3695pp; English.

XX

CC The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity

XX

SQ Sequence 459 BP; 148 A; 101 C; 102 G; 106 T; 0 U; 2 Other;

Query Match

2.2%; Score 35.2; DB 4; Length 459;

Best Local Similarity 62.5%; Pred. No. 4.2;
Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy      1451 CCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAGTCCTGCCTAA 1510
          | | | | |         |   |   | | | | | | | | | | | | | | | |
Db      250 CTTAATACCTCCAGCAACCAGTTGTGACAATACATGCAAAGAGTGCAAAGTCTTGTCCAC 309

Qy      1511 CACGTAAGTGCCTTCTTTGCTTGTTTGA 1538
          | | | | | | | | | | | | | |
Db      310 GACGGATGTTCTTTTTTTTTTTTTTTTGA 337
```

RESULT 45

AAZ16007/c

ID AAZ16007 standard; cDNA; 760 BP.

XX

AC AAZ16007;

XX

DT 12-OCT-1999 (first entry)

XX

DE Human gene expression product cDNA sequence SEQ ID NO:3476.

XX

KW Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX

OS Homo sapiens.

XX

PN WO9938972-A2.

XX

PD 05-AUG-1999.

XX

PF 28-JAN-1999; 99WO-US001619.

XX

PR 28-JAN-1998; 98US-0072910P.

PR 24-FEB-1998; 98US-0075954P.

PR 31-MAR-1998; 98US-0080114P.

PR 03-APR-1998; 98US-0080515P.

PR 03-APR-1998; 98US-0080666P.

PR 21-OCT-1998; 98US-0105234P.

PR 28-OCT-1998; 98US-0105877P.

XX

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;

XX

DR WPI; 1999-494092/41.

XX

PT Novel human genes and their expression products which are differentially

PT expressed in different cell types.

XX

PS Claim 1; Page 1661-1662; 2479pp; English.

XX

CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one of
 CC the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists

XX

SQ Sequence 760 BP; 169 A; 210 C; 174 G; 181 T; 0 U; 26 Other;

Query Match 2.2%; Score 35.2; DB 2; Length 760;

Best Local Similarity 52.6%; Pred. No. 5.6;

Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 760 AAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTTGACTCGCCACCACCT 819
 ||||| ||| | | | | ||| | || || | || | || ||
 Db 705 AAGCCAGGGGCTCCNTTTTAATTCAATTCAGGGGGTGGGTTTTTTNAAACGCAGGGCAACT 646
 Qy 820 GTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAATGGAATGA 879
 | | | ||| || ||| ||||| ||| ||| ||| |||| | | ||
 Db 645 TTTTATATAAANTCGAGGGTGCCAGGAAAGTGGGCCTGCNGGGTGCANAAAAGCGCAAGA 586
 Qy 880 AACTGCTGAAGG 892
 | || ||| |
 Db 585 AGCTTGTGGAATG 573

RESULT 46

AAS68011/c

ID AAS68011 standard; cDNA; 2412 BP.

XX

AC AAS68011;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #3815.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG03824.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 3815; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2412 BP; 537 A; 743 C; 734 G; 398 T; 0 U; 0 Other;

Query Match 2.2%; Score 35.2; DB 5; Length 2412;
Best Local Similarity 55.8%; Pred. No. 11;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 57 CAGCCATGACCA GTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTC 116
| | | | | | | | | | | | | | | | | | | | | |
Db 1797 CAGCAGTGCCCGCTCCATTTCCGGCCTCTCGGGCTGACTCCTGCAGCTGCTGCTCTAGCTC 1738
Qy 117 TGTCTTTGCTCCTTAGAGCTGGGGACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAG 176
| | | | | | | | | | | | | | | | | | | | | |
Db 1737 CTTACACGGACCTTGAGCTGCTCCACGTGCCCCAGCACCTGAGCCCGCTCTTCTCCAG 1678

RESULT 47

ADC32278/c

ID ADC32278 standard; cDNA; 2412 BP.

XX

AC ADC32278;

XX

DT 18-DEC-2003 (first entry)

XX

DE Human novel cDNA contig sequence, SEQ ID NO:2360.

XX

KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 11q23; ss.

XX

OS Homo sapiens.

XX

PN WO2003029271-A2.

XX

PD 10-APR-2003.

XX

PF 24-SEP-2002; 2002WO-US030474.

XX

PR 24-SEP-2001; 2001US-0324631P.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;

PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;

PI Haley-Vicente D, Drmanac RT;

XX

DR WPI; 2003-371981/35.

DR

P-PSDB; ADC33045.

XX

PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.

XX

PS Example 2; SEQ ID NO 2360; 1185pp; English.

XX

CC The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or

CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a human contig
 CC sequence used in an example of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 2412 BP; 537 A; 743 C; 734 G; 398 T; 0 U; 0 Other;

Query Match 2.2%; Score 35.2; DB 9; Length 2412;
 Best Local Similarity 55.8%; Pred. No. 11;
 Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 57 CAGCCATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTC 116
 |||| || || | | || | || | || | | |||| || || ||
 Db 1797 CAGCAGTGCCCGCTCCATTTCGGCCTCTCGGGCTGACTCCTGCAGCTGCTGCTCTAGCTC 1738
 Qy 117 TGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAG 176
 || | | |||| | || || | | || || || || || || || ||
 Db 1737 CTTACACGGACCTTGAGCTGCTCCACGTGCCCCAGCACCTGAGCCCGCTCTTCTTCCAG 1678

RESULT 48

ABN59902/c

ID ABN59902 standard; cDNA; 4866 BP.

XX

AC ABN59902;

XX

DT 28-JUN-2002 (first entry)

XX

DE Novel human coding sequence SEQ ID NO: 313.

XX

KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;

KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;

KW neuroprotective; antiparkinsonian; protein therapy; EST;

KW expressed sequence tag; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200222660-A2.

XX

PD 21-MAR-2002.
 XX
 PF 10-SEP-2001; 2001WO-US026015.
 XX
 PR 11-SEP-2000; 2000US-00659671.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 DR WPI; 2002-292408/33.
 DR P-PSDB; ABB97489.
 XX
 PT An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis.
 XX
 PS Claim 1; SEQ ID NO 313; 509pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a coding sequence of the
 CC invention
 XX
 SQ Sequence 4866 BP; 1038 A; 1481 C; 1463 G; 884 T; 0 U; 0 Other;

 Query Match 2.2%; Score 35.2; DB 6; Length 4866;
 Best Local Similarity 55.8%; Pred. No. 17;
 Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

 Qy 57 CAGCCATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTC 116
 |||| || || | | || | || | | |||| || || ||
 Db 1838 CAGCAGTGCCCGCTCCATTTCGGCCTCTCGGGCTGACTCCTGCAGCTGCTGCTCTAGCTC 1779

 Qy 117 TGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAG 176
 || | | |||| || || || | |||| || || |||| ||||
 Db 1778 CTTCACACGGACCTTGAGCTGCTCCACGTGCCCCAGCACCTGAGCCCGCTCTTCTTCCAG 1719

 RESULT 49
 AAS72352/c
 ID AAS72352 standard; cDNA; 5011 BP.
 XX
 AC AAS72352;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #8156.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI; 2001-639362/73.
DR P-PSDB; ABG08165.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX

PS Claim 1; SEQ ID NO 8156; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 5011 BP; 1096 A; 1500 C; 1500 G; 915 T; 0 U; 0 Other;

Query Match 2.2%; Score 35.2; DB 5; Length 5011;
Best Local Similarity 55.8%; Pred. No. 17;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 57 CAGCCATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTC 116
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1841 CAGCAGTGCCCGCTCCATTTCGGCCTCTCGGGCTGACTCCTGCAGCTGCTGCTCTAGCTC 1782

Qy 117 TGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAG 176

Db 1781 CTTACACGGACCTTGAGCTGCTCCACGTGCCCCAGCACCTGAGCCCGCTCTTCTTCCAG 1722

RESULT 50

ABK33144/c

ID ABK33144 standard; DNA; 907 BP.

XX

AC ABK33144;

XX

DT 08-MAY-2002 (first entry)

XX

DE DNA encoding novel secreted protein Z931276G1P.

XX

KW Protein secretion; mammalian secreted polypeptide; MSP; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200202621-A2.

XX

PD 10-JAN-2002.

XX

PF 28-JUN-2001; 2001WO-US020638.

XX

PR 30-JUN-2000; 2000US-0215446P.

XX

PA (ZYMO) ZYMOGENETICS INC.

XX

PI Sheppard PO, Presnell SR;

XX

DR WPI; 2002-147999/19.

DR

P-PSDB; AAU83229.

XX

PT Novel isolated mammalian secreted polypeptide useful in therapeutic and
PT diagnostic methods, to direct secretion of other proteins of interest
PT from host cell, as educational tools, and as laboratory practicum kits.

XX

PS Claim 3; Page 374-376; 397pp; English.

XX

CC The invention describes an isolated mammalian secreted polypeptide (MSP)
CC (I). (I) is useful to direct the secretion of other proteins of interest
CC from a host cell, to monitor secretion of proteins, to degenerate
CC sequences comprising all nucleotide sequences encoding a particular
CC polypeptide, to screen for cell metabolism effecting receptors, for
CC identifying new target receptors and drug design, for identifying, for
CC protein purification, for determining the weight of expressed MSP
CC polypeptides as a ratio to total protein expressed, for identifying
CC peptide cleavage sites, for coupling amino and carboxy terminal tags, for
CC amino acid sequence analysis, for monitoring biological activities of the
CC protein in vitro and in vivo, and to teach analytical skills and as
CC reagents for the study of cells, receptors, and other binding molecules.
CC The polynucleotide is useful for radiation hybrid mapping, and somatic
CC cell genetic technique developed for constructing high-resolution,
CC contiguous maps of mammalian chromosomes. Reagents disclosed in the

CC invention may be used to detect metabolic abnormalities characterised by
CC over or under production of the protein. This sequence encodes an
CC mammalian secreted polypeptide, described in the method of the invention
XX
SQ Sequence 907 BP; 199 A; 299 C; 271 G; 138 T; 0 U; 0 Other;

Query Match 2.2%; Score 35; DB 6; Length 907;
Best Local Similarity 47.1%; Pred. No. 7.2;
Matches 107; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```
Qy      706 CTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCG 765
      ||| ||| || | | | || | | | | | || | | ||| || |||
Db      468 CTCTGCCTTCTTCATCAGGGAGCTGGGACCTCGGACCAAGGCTCGGCCTTGGTGCAGCCT 409

Qy      766 CTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCTG 825
      | ||| || | | || |||| | || | || | || | || | |
Db      408 CAACCTGTCCCTGCGTCTCTTGTGCTCCCTGCTGCGTTTCCTGGTCCTCTCCCAACCTGG 349

Qy      826 TGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
      | |||| | || || | || | || | ||
Db      348 AGGTCGGTTCTCCTTCTCGGGGTAGGGCAATGCCAGTCCCAGGAGCAGGTCCTTGTCTG 289

Qy      886 CTGAAGGAATGCAGGGTTCACCTTCAAGAAGAAAGCAGTGTGCAGGTG 932
      | | ||| ||||| || ||| | || | | || ||
Db      288 CAGCAGGCCTGCAGGACTCTGCTCAGGCAGCAGCCCCTCAGCCTCTG 242
```

Search completed: April 29, 2004, 15:06:56
Job time : 761.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 14:53:14 ; Search time 143.666 Seconds
(without alignments)
6064.561 Million cell updates/sec

Title: US-09-989-981A-9_COPY_3436_5005
Perfect score: 1570
Sequence: 1 cgaagcatcctgaagtacag.....ctagagagcaaaccagagc 1570

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		8					
Result		Query					
No.	Score	Match	Length	DB	ID	Description	
1	150.8	9.6	235	3	US-09-172-108-8	Sequence 8, Appli	
2	48.6	3.1	7218	1	US-08-232-463-14	Sequence 14, Appl	
c 3	47.2	3.0	7218	1	US-08-232-463-14	Sequence 14, Appl	
c 4	39.2	2.5	289	3	US-09-007-005-17	Sequence 17, Appl	
c 5	39.2	2.5	289	3	US-09-244-796-17	Sequence 17, Appl	
6	38.4	2.4	832	4	US-09-621-976-2813	Sequence 2813, Ap	
c 7	36.2	2.3	2236	4	US-09-389-956-11	Sequence 11, Appl	
c 8	36.2	2.3	2488	4	US-09-389-956-9	Sequence 9, Appli	
c 9	35.6	2.3	399	4	US-09-621-976-8976	Sequence 8976, Ap	
10	33.6	2.1	1472	4	US-09-540-224-3	Sequence 3, Appli	
11	33.6	2.1	1472	4	US-09-564-595D-52	Sequence 52, Appl	

	12	33.6	2.1	1472	4	US-09-808-972-3	Sequence 3, Appli
c	13	33.6	2.1	1651	4	US-09-484-970B-17	Sequence 17, Appl
	14	33.2	2.1	364	4	US-09-621-976-17202	Sequence 17202, A
c	15	33.2	2.1	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	16	33.2	2.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	17	33.2	2.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	18	33	2.1	80246	3	US-09-078-294-4	Sequence 4, Appli
	19	33	2.1	80595	3	US-09-078-294-3	Sequence 3, Appli
c	20	32.8	2.1	1119	4	US-10-162-012-41	Sequence 41, Appl
c	21	32.8	2.1	1630	4	US-10-162-012-39	Sequence 39, Appl
c	22	32.8	2.1	4346	4	US-09-064-199-12	Sequence 12, Appl
c	23	32.8	2.1	4366	4	US-09-064-199-14	Sequence 14, Appl
c	24	32.8	2.1	4418	4	US-09-064-199-13	Sequence 13, Appl
c	25	32.8	2.1	4431	4	US-09-064-199-8	Sequence 8, Appli
c	26	32.8	2.1	4441	4	US-09-641-999-2	Sequence 2, Appli
c	27	32.8	2.1	4441	4	US-09-064-199-10	Sequence 10, Appl
c	28	32.8	2.1	4543	2	US-08-519-547A-5	Sequence 5, Appli
c	29	32.8	2.1	4549	4	US-09-064-199-9	Sequence 9, Appli
c	30	32.8	2.1	4564	4	US-09-064-199-2	Sequence 2, Appli
c	31	32.8	2.1	4649	4	US-09-064-199-11	Sequence 11, Appl
c	32	32.8	2.1	4746	4	US-09-064-199-7	Sequence 7, Appli
c	33	32.8	2.1	5105	4	US-09-064-199-3	Sequence 3, Appli
c	34	32.8	2.1	5463	4	US-09-064-199-1	Sequence 1, Appli
	35	32.2	2.1	648	4	US-09-489-039A-5645	Sequence 5645, Ap
c	36	32.2	2.1	831	4	US-09-489-039A-5850	Sequence 5850, Ap
	37	32.2	2.1	5683	4	US-09-845-583A-7	Sequence 7, Appli
	38	32.2	2.1	5702	4	US-09-566-921-118	Sequence 118, App
	39	31.8	2.0	75395	4	US-09-984-890-3	Sequence 3, Appli
c	40	31.6	2.0	3675	4	US-09-930-872-3	Sequence 3, Appli
c	41	31.6	2.0	4042	4	US-09-930-872-5	Sequence 5, Appli
c	42	31.6	2.0	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	43	31.4	2.0	1626	3	US-08-959-381A-4	Sequence 4, Appli
c	44	31.2	2.0	441	3	US-09-060-756-352	Sequence 352, App
c	45	31.2	2.0	441	4	US-09-670-314-352	Sequence 352, App
c	46	31.2	2.0	459	4	US-09-621-976-1509	Sequence 1509, Ap
c	47	30.8	2.0	505	4	US-09-621-976-15639	Sequence 15639, A
	48	30.8	2.0	1207	4	US-09-219-194-1	Sequence 1, Appli
c	49	30.8	2.0	2676	4	US-09-489-039A-4738	Sequence 4738, Ap
	50	30.8	2.0	2919	4	US-09-489-039A-4696	Sequence 4696, Ap

ALIGNMENTS

RESULT 1

US-09-172-108-8

; Sequence 8, Application US/09172108

; Patent No. 6160104

; GENERAL INFORMATION:

; APPLICANT: Cunnigham, Mary Jane

; APPLICANT: Zweiger, Gary B.

; APPLICANT: Panzer, Scott R.

; APPLICANT: Seilhamer, Jeffrey J.

; TITLE OF INVENTION: MARKERS FOR PEROXISOMAL PROLIFERATORS

; FILE REFERENCE: PA-0012 US

; CURRENT APPLICATION NUMBER: US/09/172,108

; CURRENT FILING DATE: 1998-10-13


```

; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

```

```

Query Match          3.1%; Score 48.6; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 0.00019;
Matches 10; Conservative 143; Mismatches 78; Indels 0; Gaps 0;

```

```

Qy      1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTCTCAGC 60
      ||| | | | | | : : : : : : : : : : : : : : : : : : : :
Db      1051 CGAGGGAGCTTGCATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1110

Qy      61 CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTC 120
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      1111 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1170

Qy      121 TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCAT 180
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      1171 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1230

Qy      181 CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTC 231
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      1231 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1281

```

RESULT 3

US-08-232-463-14/c

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

RESULT 4

US-09-007-005-17/c

```
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17
```

```
Query Match          2.5%; Score 39.2; DB 3; Length 289;
Best Local Similarity 5.4%; Pred. No. 0.027;
Matches 11; Conservative 92; Mismatches 99; Indels 0; Gaps 0;
```

```
Qy      31 CAGCTGGGTCTCTTCTTTGGTTTTCTCAGCCATGACCAGTGCTGTTTGTGCCCTTTGTGT 90
          :| |:| : : : |: : : : : : : : : : : : : : : : : :
Db      237 YAYCYGYCYAYGYCYTYGYSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYN 178

Qy      91 GGCCTCCCCTGCTGTTGGGCTCTCTCTGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCC 150
          : : : : : : : : : : : : : : : : : : : : : : : :
Db      177 YSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYN 118

Qy     151 CTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCTYTCTGGCAAACACTTCCTATAAACACA 210
          : : : : : : : : : : : : : : : : : : : : : : : :
Db     117 YSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYN 58

Qy     211 CCGTGTGTTCTGCCTATTGTCG 232
          :| | |:| : : : |:| |:|
Db      57 YCYAYTYTYGYTYAYAYTYTYG 36
```

RESULT 5

US-09-244-796-17/c

```
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
```

```

; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

```

```

Query Match          2.5%; Score 39.2; DB 3; Length 289;
Best Local Similarity 5.4%; Pred. No. 0.027;
Matches 11; Conservative 92; Mismatches 99; Indels 0; Gaps 0;

```

```

Qy      31 CAGCTGGGTCTCTTCTTTGGTTTTCTCAGCCATGACCAGTGCTGTTTGTGCCCTTTGTGT 90
        :| |:| : : : |: : : : : : : : : : : : : : : : : :
Db      237 YAYCYGYCYAYGYCYTYGYSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYN 178

Qy      91 GGCCTCCCCTGCTGTTGGGCTCTCTGTCTTTGCTCCTTAGAGCTGGGGCACCTGAGCC 150
        : : : : : : : : : : : : : : : : : : : : : : :
Db      177 YSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYN 118

Qy     151 CTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCTYTCTGGCAAACACTTCCTATAAACACA 210
        : : : : : : : : : : : : : : : : : : : : : : :
Db     117 YSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYN 58

Qy     211 CCGTGTGTTCTGCCTATTGTCG 232
        :| | |: : : |:| |:|
Db      57 YCYAYTYTYGYTYAYAYTYTYG 36

```

RESULT 6

```

US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.

```

```

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

```

```

Query Match          2.4%; Score 38.4; DB 4; Length 832;
Best Local Similarity 9.7%; Pred. No. 0.087;
Matches 30; Conservative 146; Mismatches 132; Indels 0; Gaps 0;

```

```

Qy      1261 ATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGSCSGGGGCTCCTGTACTTC 1320
          :::::  :: : : :  ::  :: : : | : :: : :  : | | : ::
Db      2    RWYWWKYTTWYAKCWTKWKWSWSYWMYWKWYMKTYWRWRRKKKKAWWKYWKWTWWYWR 61

Qy      1321 TAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTA 1380
          :|: |  :: : : : : : ::| :::: : ::::: : :|:|:|  : : : :
Db      62    YAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYAMMWTRTWTGYAYYRSMMYWWRY 121

Qy      1381 AGTTGTAGAGAGGCAGCCATGCATTTGGCATTTGAATACAATCTGGTGACTTGTCTGGCT 1440
          : ::::| :  | :: : : : : : :|: : : : :  |: : : : : : : :
Db      122   RCWKKKAYYRKTTCTYSSKGWTWWKRWKKAWTTWWWKKTYWAAATRYWMMCWTKRWRASW 181

Qy      1441 GCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAAATCCCTGGAAAGAGTGGAAAG 1500
          :|::: :|:  :  |: : | | : : |: : :::: : : :  || : :
Db      182   WYCWWWGKARKWSTWRKRSYASARSKRCCYSCSWGAMSWKYMWRMWRWRGWATGAGMK 241

Qy      1501 TCCTGCCTAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCA 1560
          :|: : :  |:  : : : :|::: :|:| :  : : :  |: : :|: : |
Db      242   AWRASCMMRRKYAGSKTSYKSMWMCWTRSWKYCYTKARWTGYCYRKGGMWGKRGRWYA 301

Qy      1561 AACCCAGA 1568
          :::::
Db      302   SKKYMWKR 309

```

RESULT 7

```

US-09-389-956-11/c
; Sequence 11, Application US/09389956
; Patent No. 6586579
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,
; TITLE OF INVENTION: Antibodies and Methods
; FILE REFERENCE: P-LJ 3611
; CURRENT APPLICATION NUMBER: US/09/389,956
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 11
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1455)
US-09-389-956-11

Query Match 2.3%; Score 36.2; DB 4; Length 2236;
Best Local Similarity 50.3%; Pred. No. 0.74;
Matches 83; Conservative 2; Mismatches 80; Indels 0; Gaps 0;

```
Qy      1174 GGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGAT 1233
          | ||||| |||| | || | || | || || |||| | |
Db      713 GTGCCTTGCTGGATCCTCTGCGCCGCGCAGATGCCGTAGGCCAGGCCGGGCACAGTACTG 654

Qy      1234 GTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGC 1293
          || | || | | | || | | || || || ||||| || | ||
Db      653 GTGCAGAGGCACACCTCGCGAGGCAGGTCCCGCAGCCACTCCGGCAGCTCCGGCGGGGGC 594

Qy      1294 CTGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAG 1338
          || || : : | ||| || | | || || | |||
Db      593 GCGGCGGCCGCGAGCGCTGCTGGTGCCACAAGCCGGCGCAGCGAG 549
```

RESULT 8

US-09-389-956-9/c

; Sequence 9, Application US/09389956
; Patent No. 6586579
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,
; TITLE OF INVENTION: Antibodies and Methods
; FILE REFERENCE: P-LJ 3611
; CURRENT APPLICATION NUMBER: US/09/389,956
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2488
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1707)
US-09-389-956-9

Query Match 2.3%; Score 36.2; DB 4; Length 2488;
Best Local Similarity 50.3%; Pred. No. 0.79;
Matches 83; Conservative 2; Mismatches 80; Indels 0; Gaps 0;

```
Qy      1174 GGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGAT 1233
          | ||||| |||| | || | || | || || |||| | |
Db      713 GTGCCTTGCTGGATCCTCTGCGCCGCGCAGATGCCGTAGGCCAGGCCGGGCACAGTACTG 654

Qy      1234 GTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGC 1293
```

```

      ||      | ||| | | | ||| |      || || ||||| || | ||
Db      653 GTGCAGAGGCACACCTCGCGAGGCAGGTCCCGCAGCCACTCCGGCAGCTCCGGCGGGGGC 594

Qy      1294 CTGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAG 1338
      || || : : | ||| ||| | | || ||| | | |||
Db      593 GCGGCGGCCGAGCGCTGCTGGTGCCACAAGCCGGCGCAGCGAG 549

```

RESULT 9

```

US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8976

```

```

Query Match          2.3%; Score 35.6; DB 4; Length 399;
Best Local Similarity 11.2%; Pred. No. 0.43;
Matches 29; Conservative 121; Mismatches 110; Indels 0; Gaps 0;

```

```

Qy      540 GGTGTCCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGC 599
      :| :|::      :| : :: : : : ||: : :: : : : ::::
Db      294 KGGSTYMAMRSRRGSTGRWSYRRAMWRGSKSWGGSYYRMAGYRSSRWSWYSAMWRKKK 235

Qy      600 TCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTGTGTCTGTCCAG 659
      :|: : | : : ::::: : : :::: : : : :|::||:: : :
Db      234 MTCWKGRSSWGSRSSTGYAWMYKKSWCTSRKWMYYKKRRKKWRRKCTSTKRTCYRGSTYK 175

Qy      660 CAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTCCT 719
      |::|:: :: : : : | : :: : : : : : : : | | ||:|: ||
Db      174 CWKAYYTKRRKRWTRWTTYYYKSYMSMKKTWRMKTAYYWTKRWKMTTRTKWTWCTMCWKCT 115

Qy      720 GCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCA 779
      : :|| : : : : : | : : : :| : : | : : : : : : : :
Db      114 TYWMAGTMMYRYRRYWYYAKRAKWSKRCTWSTTCYCMKYMAKKCWSYWWSMSMMKWGKSM 55

Qy      780 CTGATTTCTGCTCTCCCTT 799
      ::|::: : : : :|:
Db      54 WWKWTYYYYYYMMKWSKMTY 35

```

RESULT 10

```

US-09-540-224-3
; Sequence 3, Application US/09540224
; Patent No. 6468543

```

```
; GENERAL INFORMATION:
; APPLICANT: Gilbertson, Debra G.
; APPLICANT: Hart, Charles E.
; TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,
; TITLE OF INVENTION: LIGAMENT AND CARTILAGE USING ZVEGF4
; FILE REFERENCE: 00-28
; CURRENT APPLICATION NUMBER: US/09/540,224
; CURRENT FILING DATE: 2000-03-31
; EARLIER APPLICATION NUMBER: US 60/180,169
; EARLIER FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1472
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(1205)
US-09-540-224-3
```

```
Query Match          2.1%; Score 33.6; DB 4; Length 1472;
Best Local Similarity 59.4%; Pred. No. 3.8;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
Qy      827 GTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAATGGAATGAACACTGC 886
          | || || ||||| ||| ||||| |  || ||  |  || ||||  || |||
Db      435 GAAGTTGAAGAAGTCTCAGAGAGCAGCACTGTTGTTCAGAGGAAGATGGTGTGGCCACAAG 494

Qy      887 TGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAG 922
          |  | || || | || ||||| ||| |||
Db      495 GAGATCCCTCCAAGGATAACGTCAAGAACAAACCAG 530
```

```
RESULT 11
US-09-564-595D-52
; Sequence 52, Application US/09564595D
; Patent No. 6495668
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Teresa
; APPLICANT: Hart, Charles E.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF4
; FILE REFERENCE: 99-19
; CURRENT APPLICATION NUMBER: US/09/564,595D
; CURRENT FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: US 09/304,216
; PRIOR FILING DATE: 1999-05-03
; PRIOR APPLICATION NUMBER: US 60/164,463
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: US 60/180,169
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 1472
; TYPE: DNA
```

; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(1205)
US-09-564-595D-52

Query Match 2.1%; Score 33.6; DB 4; Length 1472;
Best Local Similarity 59.4%; Pred. No. 3.8;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 827 GTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGC 886
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 435 GAAGTTGAAGAAGTCTCAGAGAGCAGCACTGTTGTCAGAGGAAGATGGTGTGGCCACAAG 494

Qy 887 TGAAGGAATGCAGGGTTCACCTTCAAGAAGAAAGCAG 922
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 495 GAGATCCCTCCAAGGATAACGTCAAGAACAAACCAG 530

RESULT 12

US-09-808-972-3

; Sequence 3, Application US/09808972
; Patent No. 6630142
; GENERAL INFORMATION:
; APPLICANT: Hart, Charles E.
; APPLICANT: Topouzis, Stavros
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: METHOD OF TREATING FIBROPROLIFERATIVE
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 00-79
; CURRENT APPLICATION NUMBER: US/09/808,972
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US 60/235,295
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/564,595
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: US 60/180,169
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/164,463
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: US 60/132,250
; PRIOR FILING DATE: 1999-05-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1472
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(1205)
US-09-808-972-3

Query Match 2.1%; Score 33.6; DB 4; Length 1472;
Best Local Similarity 59.4%; Pred. No. 3.8;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 827 GTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGC 886
 | | | | | | | | | | | | | | | | | | | | | |
 Db 435 GAAGTTGAAGAAGTCTCAGAGAGCAGCACTGTTGTCAGAGGAAGATGGTGTGGCCACAAG 494
 QY 887 TGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAG 922
 | | | | | | | | | | | | | | | | | |
 Db 495 GAGATCCCTCCAAGGATAACGTCAAGAACAAACCAG 530

RESULT 13

US-09-484-970B-17/c
 ; Sequence 17, Application US/09484970B
 ; Patent No. 6426186
 ; GENERAL INFORMATION:
 ; APPLICANT: Jones, Karen A.
 ; APPLICANT: Volkmuth, Wayne
 ; APPLICANT: Walker, Michael G.
 ; TITLE OF INVENTION: BONE REMODELING GENES
 ; FILE REFERENCE: PB-0014 US
 ; CURRENT APPLICATION NUMBER: US/09/484,970B
 ; CURRENT FILING DATE: 2000-01-18
 ; NUMBER OF SEQ ID NOS: 172
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 17
 ; LENGTH: 1651
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6426186 126510.2CB1
 ; NAME/KEY: unsure
 ; LOCATION: 767-846
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-484-970B-17

Query Match 2.1%; Score 33.6; DB 4; Length 1651;
 Best Local Similarity 52.9%; Pred. No. 4.1;
 Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 415 TGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCA 474
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1075 TGCCAGCAACAGACTCTCCTCCCTTGCTGAGACCAGAAGGTGAGTGAGGGCTTTGCAATG 1016
 QY 475 ACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACA 534
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1015 AGGGCAGGGCATGGAGGTGACTGTCACTCTTTGCTGGCAGGGGGTCTCAGGACTATAGGA 956
 QY 535 GCTTAGGTGTCCTGCA 550
 | | | | | | | | | | | | | | | | | |
 Db 955 ACTTTAGAGCCTTGCA 940

RESULT 14

US-09-621-976-17202
 ; Sequence 17202, Application US/09621976
 ; Patent No. 6639063
 ; GENERAL INFORMATION:


```
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 17202
; LENGTH: 364
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-17202
```

```
Query Match          2.1%; Score 33.2; DB 4; Length 364;
Best Local Similarity 13.5%; Pred. No. 2.3;
Matches 38; Conservative 118; Mismatches 126; Indels 0; Gaps 0;
```

```
Qy      146 GAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCTYTCTGGCAAACACTTCCTATAA 205
      |||:: | : ::| :|| : : | : : :| : :: : : : : : : :
Db      21 GAGYSGMCKSSRSYGRSSCCGSMGWSGCSCSKRSWSRCRCMKSMWSWMMYMRSMKYKRS 80

Qy      206 ACACACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGA 265
      ||: : : |: : |: :| :| : : : :| :| : : : : : : : :
Db      81 TCASCKYKGGKMACMTCWSTGAMYRYMASYGWCYSYMARYYTCYSKYRMWKYCYRKYRSR 140

Qy      266 TAATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTA 325
      : || :| : :| : ::| : : : : : : : : : : : : :|
Db      141 GMCCMWCAAGSGMCYSRSAGSRYSKKGSRGRWYWKKGCSRATSKKGRMMWMKKGSRRRATS 200

Qy      326 GGATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGC 385
      : : :: : : : :| :| : : : : : : :| : : : :| : :
Db      201 RYGMSSMYGASKRMSSMCSASTRMSSASCMYMMMSAGSYASCWKMSKYRRCAKWSCT 260

Qy      386 CACTAGAAAATTCACTTGCAATTTGCTTCCTGCTAGCCATGGG 427
      : : ::| : : | : : : | : : :| : : :| :| :
Db      261 YSWYMRASMKSKSKYCAWSRKGSKCCMYSRKGSKSCYCCWGGG 302
```

RESULT 15

US-09-621-976-2813/c

```
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 2.1%; Score 33.2; DB 4; Length 832;
Best Local Similarity 12.3%; Pred. No. 3.7;
Matches 20; Conservative 82; Mismatches 60; Indels 0; Gaps 0;

Qy 1340 CTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTAAGTTGTAGAGAGGCAGCCA 1399
|:::|: :::: :: |::| :::: ||::: :: ::| ::: : : :: :| | : :
Db 188 CWWWGRWWSTYWYMAWGKKWWRYATTWRRAMWWWWAAWTMMWYMWAWCMSSRGAAMYRR 129
Qy 1400 TGCATTTGGCATTGGAATACAATCTGGTGACTTGTCTGGCTGCCAATAGAACCTAGTACC 1459
| :::: :: : :::: : |::: : : : :::: :: |::: : :| : : :|:::
Db 128 TMMMWGYRYWWRKKSYYRTRCAWAYAWKTKRSYYWCWRWKWKRCMMMMMAMAYGKTMMM 69
Qy 1460 AAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAGT 1501
:| : : : : | ::::|::: : ::| :::::
Db 68 RACWKTRYWRWWAWAMWRMWWTMMMYWYWRAMKRRWMWRK 27

RESULT 16

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.1%; Score 33.2; DB 3; Length 4403765;
Best Local Similarity 52.1%; Pred. No. 1.2e+02;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 599 CTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCCTGTGGGCTTTGGGTTGTCTGTCCA 658
| | | | | | | | | | | | | | | | | | | | | |
Db 866529 CACGCAGCTCGTCTATGGTGAGTTCGTCGTTATGCAGGACGGCGTTGACTTGGACCTCGA
866470

Qy 659 GCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTCC 718
 ||||| | | | | | | | | | | | | | | | | | |
 Db 866469 GCAGATCCGACTGGCCTTGAGCGGTGACGACCGCGATCACGGCCATCCGCCGGTCCCGCA
 866410

Qy 719 TGC GCAGGGCAGAGCCTGGACA 740
 || |||| | | | | |
 Db 866409 TGGACAGGCCGGGACGGGTCCA 866388

RESULT 17

US-09-103-840A-1/c
 ; Sequence 1, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 4411529
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; OTHER INFORMATION: H37Rv
 US-09-103-840A-1

Query Match 2.1%; Score 33.2; DB 3; Length 4411529;
 Best Local Similarity 52.1%; Pred. No. 1.2e+02;
 Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 599 CTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCCTGTGGGCTTTGGGTTGTCTGTCCA 658
 | | | | | | | | | | | | | | | | | | | | | |
 Db 864399 CACGCAGCTCGTCTATGGTGAGTTCGTCTGTTATGCAGGACGGCGTTGACTTGGACCTCGA
 864340

Qy 659 GCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTCC 718
 ||||| | | | | | | | | | | | | | | | | | |
 Db 864339 GCAGATCCGACTGGCCTTGAGCGGTGACGACCGCGATCACGGCCATCCGCCGGTCCCGCA
 864280

Qy 719 TGC GCAGGGCAGAGCCTGGACA 740
 || |||| | | | | |
 Db 864279 TGGACAGGCCGGGACGGGTCCA 864258

RESULT 18

US-09-078-294-4
 ; Sequence 4, Application US/09078294
 ; Patent No. 6265211

```
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4
```

```
Query Match          2.1%; Score 33; DB 3; Length 80246;
Best Local Similarity 48.6%; Pred. No. 56;
Matches 87; Conservative 1; Mismatches 91; Indels 0; Gaps 0;
```

```
Qy          9 CCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTCTCAGCCATGACCA 68
      |||  | | ||| | ||  |  || || |  ||| | ||  ||
Db          21544 CCTCCCTTCCCTTCCCCTCCCCTCCCCTTCCCTTCTCCCTCTCCTTCCCTTCCCTCTTCCC 21603

Qy          69 GTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTGTCTTTGCTCC 128
      | ||  | | |||| | |  ||||| || | ||  | | |||
Db          21604 TTCCTTCCCTTCCCTTCCCTTCCCCTCCCCTTCCCTTCCCTTCCCTCCCTCCCTTCCCTCC 21663

Qy          129 TTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCTYTCT 187
      |  |  | |||| | ||  |||| || ||  | ||||: ||
Db          21664 CTTCTTTCCTTCCCTTCTTTCCTTCCTCATTTCCTCCCTTCCTTCCTTCCTTCCTTCCT 21722
```

RESULT 19

US-09-078-294-3

```
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3
```

```
Query Match          2.1%; Score 33; DB 3; Length 80595;
Best Local Similarity 48.6%; Pred. No. 56;
Matches 87; Conservative 1; Mismatches 91; Indels 0; Gaps 0;
```

Qy 9 CCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGCCATGACCA 68
 ||| | | ||| | | | | | | | | | | | |
 Db 21806 CCTCCCTTCCCTTCCCCTCCCCTCCCCTTCCCTTCTCCCTCTCCTTCCCTTCCTCTTCCC 21865

Qy 69 GTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTGTCTTTGCTCC 128
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21866 TTCCTTCCCTTCCCTTCCCTTCCCCTCCCCTTCCCTTCCCTTCCCTCCCTCCCTTCCTCC 21925

Qy 129 TTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCTYTCT 187
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21926 CTTCTTTCCTTCCCTTCTTTCCTTCCCTCATTTCCTCCCTTCCCTTCCCTTCCCTTCCCT 21984

RESULT 20

US-10-162-012-41/c

; Sequence 41, Application US/10162012

; Patent No. 6682597

; GENERAL INFORMATION:

; APPLICANT: Curtis, Rory A.J.

; APPLICANT: Silos-Santiago, Inmaculada

; APPLICANT: Gu, Wei

; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS

; FILE REFERENCE: 10448-190001

; CURRENT APPLICATION NUMBER: US/10/162,012

; CURRENT FILING DATE: 2002-06-04

; PRIOR APPLICATION NUMBER: US 60/209,845

; PRIOR FILING DATE: 2000-06-06

; PRIOR APPLICATION NUMBER: US 09/875,321

; PRIOR FILING DATE: 2001-06-06

; PRIOR APPLICATION NUMBER: PCT/US01/18340

; PRIOR FILING DATE: 2001-06-06

; PRIOR APPLICATION NUMBER: US 60/209,257

; PRIOR FILING DATE: 2000-06-05

; PRIOR APPLICATION NUMBER: US 09/875,423

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: PCT/US01/18398

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/209,238

; PRIOR FILING DATE: 2000-06-05

; PRIOR APPLICATION NUMBER: US 09/875,363

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: PCT/US01/18247

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/227,068

; PRIOR FILING DATE: 2000-08-22

; PRIOR APPLICATION NUMBER: US 09/928,530

; PRIOR FILING DATE: 2001-08-13

; PRIOR APPLICATION NUMBER: PCT/US01/25475

; PRIOR FILING DATE: 2001-08-15

; PRIOR APPLICATION NUMBER: US 60/226,770

; PRIOR FILING DATE: 2000-08-21

; PRIOR APPLICATION NUMBER: US 09/934,421

; PRIOR FILING DATE: 2001-08-21

; PRIOR APPLICATION NUMBER: PCT/US01/26096

; PRIOR FILING DATE: 2001-08-21

; PRIOR APPLICATION NUMBER: US 60/279,281

; PRIOR FILING DATE: 2001-03-28


```

; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 1630
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (230)...(1345)
US-10-162-012-39

```

```

Query Match          2.1%; Score 32.8; DB 4; Length 1630;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 82; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```

```

Qy      887 TGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCCAGT 946
        || || || || || || || || || || || || || || || || || || || ||
Db      617 TGGTGGCTGGCCGGGAGGACATCCAGAGGGAGAAGAGGCTGATGAGCATGCTGGCAAAGT 558

Qy      947 CAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTT 1006
        ||| || | || | || | || || || || || || || || || || || || || ||
Db      557 CAGTGAGCAGGTGTGCTGCGTCAGTCATGACAGCCAAGCTGTGTGCCAGGTACCCACCAA 498

Qy      1007 GTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAA 1050
        || | | || | || || || || || || || || || || || || || || || ||
Db      497 CGACTTCTCCGATCATGAACAACAGGCAGATGGCAGAGGCTACA 454

```

RESULT 22

US-09-064-199-12/c
; Sequence 12, Application US/09064199
; Patent No. 6632604
; GENERAL INFORMATION:
; APPLICANT: MACH, Bernard
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES
; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING
THE
; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,
AND
; THEIR USE, IN PARTICULAR AS DRUGS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,199
; FILING DATE: 22-Apr-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 97-04954
; FILING DATE: 22-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 017753-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4346 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: cIita of type II
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-064-199-12

Query Match 2.1%; Score 32.8; DB 4; Length 4346;
Best Local Similarity 54.0%; Pred. No. 12;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy	1080	TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT	1139
Db	3676	TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT	3617

Qy 1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCGGGCCTTGGTGGAACATCAAATCATG 1199
 | | | | | | | | | | | | | | | | | | | | | |
 Db 3616 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3557
 Qy 1200 CCAG 1203
 | | | |
 Db 3556 CCAG 3553

RESULT 23

US-09-064-199-14/c

; Sequence 14, Application US/09064199

; Patent No. 6632604

; GENERAL INFORMATION:

; APPLICANT: MACH, Bernard

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES

; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING

THE

; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,

AND

; THEIR USE, IN PARTICULAR AS DRUGS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/064,199

; FILING DATE: 22-Apr-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 97-04954

; FILING DATE: 22-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Rea, Teresa Stanek

; REGISTRATION NUMBER: 30,427

; REFERENCE/DOCKET NUMBER: 017753-096

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4366 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

```

;          NAME/KEY:  cIIIta of type IV
;          SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-064-199-14

```

Query Match 2.1%; Score 32.8; DB 4; Length 4366;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy	1080	TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT	1139
Db	3696	TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT	3637
Qy	1140	GCCCTTGTGGGATTTCCTTTAAAGCAACCGTGTGGGCGCTTGGTGGAACATCAAATCATG	1199
Db	3636	GGCTGTCCGCAATGTCCTTCAGAGAAGGCCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA	3577
Qy	1200	CCAG 1203	
Db	3576	CCAG 3573	

RESULT 24

US-09-064-199-13/c

; Sequence 13, Application US/09064199

; Patent No. 6632604

; GENERAL INFORMATION:

APPLICANT: MACH, Bernard

: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES

WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING

THE

EXPRESSON OF GENES ENCODING MHC TYPE II MOLECULES,

AND

THEIR USE, IN PARTICULAR AS DRUGS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

```

: MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/064,199

FILING DATE: 22-Apr-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 97-04954

FILING DATE: 22-APR-1997

ATTORNEY/AGENT INFORMATION:

```

; NAME: Rea, Teresa Stanek

```

REGISTRATION NUMBER: 30,427

REFERENCE/DOCKET NUMBER: 017753-096

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4418 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: cIIIta of type III
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-064-199-13

```

```

Query Match          2.1%; Score 32.8; DB 4; Length 4418;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
          ||| | ||| |||| || || | ||| ||| | | || | ||| | |
Db      3748 TGGCAGGGGCCTGGGAGGCAGAAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 3689

Qy      1140 GCCCTTGTGGGATTTTCCTTTTAAAGCAACCGTGTCTGGGCCTTGGTGGAACATCAAATCATG 1199
          | | | | || |||| | || | | | || | ||| | || | || |
Db      3688 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3629

Qy      1200 CCAG 1203
          ||||
Db      3628 CCAG 3625

```

RESULT 25

US-09-064-199-8/c

```

; Sequence 8, Application US/09064199
; Patent No. 6632604
; GENERAL INFORMATION:
; APPLICANT: MACH, Bernard
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES
; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING
THE
; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,
AND
; THEIR USE, IN PARTICULAR AS DRUGS
;
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/064,199
;      FILING DATE: 22-Apr-1998
;      CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: FR 97-04954
;      FILING DATE: 22-APR-1997
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Rea, Teresa Stanek
;      REGISTRATION NUMBER: 30,427
;      REFERENCE/DOCKET NUMBER: 017753-096
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (703) 836-6620
;      TELEFAX: (703) 836-2021
;      INFORMATION FOR SEQ ID NO: 8:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 4431 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: DNA (genomic)
;      FEATURE:
;      NAME/KEY: cIIIta de type II
;      SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-064-199-8

```

```

Query Match          2.1%;  Score 32.8;  DB 4;  Length 4431;
Best Local Similarity 54.0%;  Pred. No. 13;
Matches 67;  Conservative 0;  Mismatches 57;  Indels 0;  Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
          ||| | |||  |||| || || | ||| ||| | | ||  ||| ||| | |
Db      3761 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 3702

Qy      1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCTGGGCCTTGGTGGAACATCAAATCATG 1199
          | | | | | || |||| | || | | | || | ||| | || | || |
Db      3701 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3642

Qy      1200 CCAG 1203
          ||||
Db      3641 CCAG 3638

```

RESULT 26

US-09-641-999-2/c

```

; Sequence 2, Application US/09641999
; Patent No. 6379894
; GENERAL INFORMATION:
; APPLICANT: MACH, BERNARD
; TITLE OF INVENTION: METHOD FOR SCREENING COMPOUNDS CAPABLE OF INHIBITING
; TITLE OF INVENTION: FIXING BETWEEN THE STAT1 TRANSCRIPTION FACTOR AND THE
; TITLE OF INVENTION: USF1 TRANSCRIPTION FACTOR
; FILE REFERENCE: EGYPT 3.3-007CONT
; CURRENT APPLICATION NUMBER: US/09/641,999
; CURRENT FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 2
; LENGTH: 4441
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-641-999-2

Query Match 2.1%; Score 32.8; DB 4; Length 4441;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
||| | ||| |||| | | | ||| ||| | | || | || |
Db 3771 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 3712

Qy 1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTGCGGCCTTGGTGGACATCAAATCATG 1199
| | | | | |||| | | | | | | | ||| | | | |
Db 3711 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3652

Qy 1200 CCAG 1203
||| |
Db 3651 CCAG 3648

RESULT 27

US-09-064-199-10/c

; Sequence 10, Application US/09064199

; Patent No. 6632604

; GENERAL INFORMATION:

; APPLICANT: MACH, Bernard

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES

; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING

THE

; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,

AND

; THEIR USE, IN PARTICULAR AS DRUGS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/064,199

; FILING DATE: 22-Apr-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 97-04954

; FILING DATE: 22-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Rea, Teresa Stanek

```

;          REGISTRATION NUMBER: 30,427
;          REFERENCE/DOCKET NUMBER: 017753-096
;          TELECOMMUNICATION INFORMATION:
;          TELEPHONE: (703) 836-6620
;          TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 10:
;      SEQUENCE CHARACTERISTICS:
;          LENGTH: 4441 base pairs
;          TYPE: nucleic acid
;          STRANDEDNESS: single
;          TOPOLOGY: linear
;          MOLECULE TYPE: DNA (genomic)
;          FEATURE:
;              NAME/KEY: cIIta of type IV
;          SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-064-199-10

```

```

Query Match          2.1%;  Score 32.8;  DB 4;  Length 4441;
Best Local Similarity 54.0%;  Pred. No. 13;
Matches 67;  Conservative 0;  Mismatches 57;  Indels 0;  Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
          ||| | |||  |||| || || | ||| ||| | | ||  ||  ||| | |
Db      3771 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 3712

Qy      1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCGGGCCTTGGTGGACATCAAATCATG 1199
          | | | | || |||| | || | | | || | ||| | || || |
Db      3711 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3652

Qy      1200 CCAG 1203
          ||||
Db      3651 CCAG 3648

```

RESULT 28

US-08-519-547A-5/c

```

; Sequence 5, Application US/08519547A
; Patent No. 5994082
;  GENERAL INFORMATION:
;      APPLICANT:
;      TITLE OF INVENTION:  Proteins Essential for the Expression of
;      TITLE OF INVENTION:  Vertebrate MHC Class II Genes, DNA Sequences Encoding
Same
;      TITLE OF INVENTION:  and Pharmaceutical Compositions
;      NUMBER OF SEQUENCES: 6
;      CORRESPONDENCE ADDRESS:
;          ADDRESSEE: FISH & NEAVE
;          STREET: 1251 AVENUE OF THE AMERICAS
;          CITY: NEW YORK
;          STATE: NEW YORK
;          COUNTRY: U.S.A.
;          ZIP: 10020-1104
;      COMPUTER READABLE FORM:
;          MEDIUM TYPE: Floppy disk
;          COMPUTER: IBM PC compatible
;          OPERATING SYSTEM: MS-DOS
;          SOFTWARE: WordPerfect 6.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/519,547A
; FILING DATE: 25-AUG-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP94113378.7
; FILING DATE: 26-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: HALEY, JAMES F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: VOS-11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-519-547A-5

```

```

Query Match          2.1%; Score 32.8; DB 2; Length 4543;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
          ||| | ||| |||| || || || || || || || || || || || || || || || ||
Db      3873 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 3814

Qy      1140 GCCCTTGTGGGATTTCTTTAAAGCAACCGTGTGCGGGCCTTGGTGGAACATCAAATCATG 1199
          || | | | || ||||| || | || | || | || | || | || | || |
Db      3813 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3754

Qy      1200 CCAG 1203
          ||||
Db      3753 CCAG 3750

```

RESULT 29

US-09-064-199-9/c

; Sequence 9, Application US/09064199

; Patent No. 6632604

; GENERAL INFORMATION:

; APPLICANT: MACH, Bernard

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES

; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING
THE

; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,

AND

; THEIR USE, IN PARTICULAR AS DRUGS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

```

; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,199
; FILING DATE: 22-Apr-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 97-04954
; FILING DATE: 22-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 017753-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4549 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: cIIIta of type III
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-064-199-9

```

```

Query Match          2.1%; Score 32.8; DB 4; Length 4549;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
          ||| | ||| |||| || || | ||| ||| | | || | ||| | |
Db      3879 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 3820

Qy      1140 GCCCTTGTGGGATTTCTTTAAAGCAACCGTGTCTGGGCCTTGGTGGAACATCAAATCATG 1199
          | | | | | |||| | || | | | || | ||| | || | |
Db      3819 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3760

Qy      1200 CCAG 1203
          |||
Db      3759 CCAG 3756

```

```

RESULT 30
US-09-064-199-2/c
; Sequence 2, Application US/09064199
; Patent No. 6632604

```



```

; GENERAL INFORMATION:
; APPLICANT: MACH, Bernard
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES
; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING
THE
; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,
AND
; THEIR USE, IN PARTICULAR AS DRUGS
;
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,199
; FILING DATE: 22-Apr-1998
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 97-04954
; FILING DATE: 22-APR-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 017753-096
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4564 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: cIIIta gene of type II
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-064-199-2

```

```

Query Match          2.1%; Score 32.8; DB 4; Length 4564;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
        ||| | |||  |||| || || | ||| ||| | | ||  ||| || |
Db      3894 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 3835

Qy      1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCTGGGCCTTGGTGGAAACATCAAATCATG 1199
        | | | | | || |||| | || | | | | || | || | || |

```

Db 3834 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3775
Qy 1200 CCAG 1203
||||
Db 3774 CCAG 3771

RESULT 31

US-09-064-199-11/c

; Sequence 11, Application US/09064199

; Patent No. 6632604

; GENERAL INFORMATION:

; APPLICANT: MACH, Bernard

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES

; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING

THE

; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,

AND

; THEIR USE, IN PARTICULAR AS DRUGS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/064,199

; FILING DATE: 22-Apr-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 97-04954

; FILING DATE: 22-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Rea, Teresa Stanek

; REGISTRATION NUMBER: 30,427

; REFERENCE/DOCKET NUMBER: 017753-096

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4649 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: cIIta of type I

; SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-09-064-199-11

Query Match 2.1%; Score 32.8; DB 4; Length 4649;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy	1080	TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT	1139
Db	3979	TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT	3920
Qy	1140	GCCCTTGTGGGATTTCCTTTAAAGCAACCGTGTCGGGCCCTTGGTGGAACATCAAATCATG	1199
Db	3919	GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA	3860
Qy	1200	CCAG 1203	
Db	3859	CCAG 3856	

RESULT 32

US-09-064-199-7/c

; Sequence 7, Application US/09064199

; Patent No. 6632604

; GENERAL INFORMATION:

APPLICANT: MACH, Bernard

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES

WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING

THE

EXPRESSON OF GENES ENCODING MHC TYPE II MOLECULES,

AND

THEIR USE, IN PARTICULAR AS DRUGS

```

;          NUMBER OF SEQUENCES: 25

```

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible

```

;      OPERATING SYSTEM: PC-DOS/MS-DOS

```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/09/064,199

; FILING DATE: 22-Apr-1998

; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 97-04954

FILING DATE: 22-APR-1997

; ATTORNEY/AGENT INFORMATION:

NAME: Rea, Teresa Stanek

REGISTRATION NUMBER: 30,427

REFERENCE/DOCKET NUMBER: 017753-096

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

```
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 4746 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;     FEATURE:
;         NAME/KEY: cIIIta of type I
;     SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-064-199-7
```

```
Query Match          2.1%; Score 32.8; DB 4; Length 4746;
Best Local Similarity 54.0%; Pred. No. 13;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
      ||| | ||| |||| || || | ||| ||| | | || | ||| | |
Db      4076 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 4017

Qy      1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTGCGGGCCTTGGTGGAACATCAAATCATG 1199
      | | | | || |||| | || | | | || | ||| | || | |
Db      4016 GGCTGTCCGCAATGTCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 3957

Qy      1200 CCAG 1203
      ||||
Db      3956 CCAG 3953
```

RESULT 33

US-09-064-199-3/c

; Sequence 3, Application US/09064199

; Patent No. 6632604

; GENERAL INFORMATION:

; APPLICANT: MACH, Bernard

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES

; WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING

THE

; EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,

AND

; THEIR USE, IN PARTICULAR AS DRUGS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/064,199

; FILING DATE: 22-Apr-1998

```

;          CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: FR 97-04954
;          FILING DATE: 22-APR-1997
;
; ATTORNEY/AGENT INFORMATION:
;          NAME: Rea, Teresa Stanek
;          REGISTRATION NUMBER: 30,427
;          REFERENCE/DOCKET NUMBER: 017753-096
;
; TELECOMMUNICATION INFORMATION:
;          TELEPHONE: (703) 836-6620
;          TELEFAX: (703) 836-2021
;
; INFORMATION FOR SEQ ID NO: 3:
;          SEQUENCE CHARACTERISTICS:
;              LENGTH: 5105 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;
;          MOLECULE TYPE: DNA (genomic)
;
;          FEATURE:
;              NAME/KEY: cIIIta gene of type IV
;
;          SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-064-199-3

```

```

Query Match          2.1%; Score 32.8; DB 4; Length 5105;
Best Local Similarity 54.0%; Pred. No. 14;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
          ||| | ||| |||| || || | ||| ||| | | || | ||| | |
Db      4435 TGGCAGGGGCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 4376

Qy      1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTGCGGCCTTGGTGGAAACATCAAATCATG 1199
          | | | | | ||||| || | | | | | ||| | | | | |
Db      4375 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 4316

Qy      1200 CCAG 1203
          ||||
Db      4315 CCAG 4312

```

RESULT 34

US-09-064-199-1/c

```

; Sequence 1, Application US/09064199
; Patent No. 6632604
;
; GENERAL INFORMATION:
;
; APPLICANT: MACH, Bernard
;
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF CIITA GENES
;                      WHICH CAN BE INVOLVED IN CONTROLLING AND REGULATING
THE
;                      EXPRESSION OF GENES ENCODING MHC TYPE II MOLECULES,
AND
;                      THEIR USE, IN PARTICULAR AS DRUGS
;
; NUMBER OF SEQUENCES: 25
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;
; STREET: P.O. Box 1404
;
; CITY: Alexandria

```

```

;          STATE: Virginia
;          COUNTRY: United States
;          ZIP: 22313-1404
;
;  COMPUTER READABLE FORM:
;          MEDIUM TYPE: Floppy disk
;          COMPUTER: IBM PC compatible
;          OPERATING SYSTEM: PC-DOS/MS-DOS
;          SOFTWARE: PatentIn Release #1.0, Version #1.30
;
;  CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/09/064,199
;          FILING DATE: 22-Apr-1998
;          CLASSIFICATION: <Unknown>
;
;  PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: FR 97-04954
;          FILING DATE: 22-APR-1997
;
;  ATTORNEY/AGENT INFORMATION:
;          NAME: Rea, Teresa Stanek
;          REGISTRATION NUMBER: 30,427
;          REFERENCE/DOCKET NUMBER: 017753-096
;
;  TELECOMMUNICATION INFORMATION:
;          TELEPHONE: (703) 836-6620
;          TELEFAX: (703) 836-2021
;
;  INFORMATION FOR SEQ ID NO: 1:
;          SEQUENCE CHARACTERISTICS:
;              LENGTH: 5463 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;
;          MOLECULE TYPE: DNA (genomic)
;          FEATURE:
;              NAME/KEY: cIIta gene of type I
;
;          SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-064-199-1

```

```

Query Match          2.1%;  Score 32.8;  DB 4;  Length 5463;
Best Local Similarity 54.0%;  Pred. No. 14;
Matches 67;  Conservative 0;  Mismatches 57;  Indels 0;  Gaps 0;

```

```

Qy      1080 TGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTT 1139
          ||| | |||  |||| || || | ||| ||| | | ||  ||  ||| | |
Db      4793 TGGCAGGGGCCTGGGCAGGCAGAATGGGGCTGCCTCTGTCACTCCCTCTGGCCTGGCCGT 4734

Qy      1140 GCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCTGGGCCTTGGTGGACATCAAATCATG 1199
          | | | | || |||| | || | | | || | ||| | || | || |
Db      4733 GGCTGTCCGCAATGTCCTTCAGAGAAGGCCTCGGGGCTCCTGGGTGGACCGCAGCTGCCA 4674

Qy      1200 CCAG 1203
          ||||
Db      4673 CCAG 4670

```

```

RESULT 35
US-09-489-039A-5645
; Sequence 5645, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 5645
 ; LENGTH: 648
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-5645

Query Match 2.1%; Score 32.2; DB 4; Length 648;
 Best Local Similarity 56.0%; Pred. No. 6.6;
 Matches 61; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 66 CCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTCTTTGC 125
 || |||| | | || | || | | || || || || || || || ||
 Db 370 CCGCTGCTGCTGGCCGCCGAGAGTTGCGCCGGCATCTGCTGCGCTTTTCTTCCAGTCC 429
 Qy 126 TCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCC 174
 | | || | || || | || | || || || || || || ||
 Db 430 GACCTGACGCAGCAGCGCCAGCGCCCGCGCCTGAGCCGCGCTTTTCCCC 478

RESULT 36

US-09-489-039A-5850/c
 ; Sequence 5850, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 5850
 ; LENGTH: 831
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-5850

Query Match 2.1%; Score 32.2; DB 4; Length 831;
 Best Local Similarity 56.0%; Pred. No. 7.6;
 Matches 61; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 66 CCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTCTTTGC 125
 || |||| | | || | || | | || || || || || || || ||
 Db 513 CCGCTGCTGCTGGCCGCCGAGAGTTGCGCCGGCATCTGCTGCGCTTTTCTTCCAGTCC 454

Qy 126 TCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCC 174
| | | | | | | | | | | | | | | | | |
Db 453 GACCTGACGCAGCAGCGCCAGCGCCCGCGCCTGAGCCGCGCTTTTCCCC 405

RESULT 37
US-09-845-583A-7
; Sequence 7, Application US/09845583A

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 17:06:46 ; Search time 1526.81 Seconds
(without alignments)
4651.434 Million cell updates/sec

Title: US-09-989-981A-9_COPY_3436_5005
Perfect score: 1570
Sequence: 1 cgaagcatcctgaagtacag.....ctagagagcaaaccagagc 1570

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result 8
Query

No.	Score	Match	Length	DB	ID	Description
1	1568	99.9	6043	10	US-09-989-981A-9	Sequence 9, Appli
2	358.6	22.8	359	10	US-09-989-981A-10	Sequence 10, Appl
3	215	13.7	2512	16	US-10-104-047-825	Sequence 825, App
4	191.4	12.2	2258	9	US-09-837-992-2	Sequence 2, Appli
5	146.4	9.3	1959	10	US-09-989-981A-1	Sequence 1, Appli
6	101.6	6.5	249	9	US-09-837-992-7	Sequence 7, Appli
7	101.6	6.5	2340	9	US-09-837-992-4	Sequence 4, Appli
8	101.6	6.5	2340	10	US-09-989-981A-5	Sequence 5, Appli
9	90	5.7	122	9	US-09-837-992-8	Sequence 8, Appli
c 10	63	4.0	2019	10	US-09-989-981A-3	Sequence 3, Appli
11	39.2	2.5	1184	13	US-10-142-426-412	Sequence 412, App
12	39.2	2.5	1184	15	US-10-123-155-412	Sequence 412, App
13	39.2	2.5	1184	15	US-10-146-731-412	Sequence 412, App
14	39.2	2.5	1184	15	US-10-140-472-412	Sequence 412, App
15	39.2	2.5	1184	15	US-10-141-761-412	Sequence 412, App
16	39.2	2.5	1184	15	US-10-142-885-412	Sequence 412, App
17	39.2	2.5	1184	15	US-10-158-790-412	Sequence 412, App
18	39.2	2.5	1184	16	US-10-137-871-412	Sequence 412, App
19	39.2	2.5	1184	16	US-10-140-923-412	Sequence 412, App
20	39.2	2.5	1184	16	US-10-141-756-412	Sequence 412, App
21	39.2	2.5	1184	16	US-10-141-759-412	Sequence 412, App
22	39.2	2.5	1184	16	US-10-140-805-412	Sequence 412, App
23	39.2	2.5	1184	16	US-10-140-864-412	Sequence 412, App
24	38.2	2.4	686	15	US-10-184-644-524	Sequence 524, App
25	38.2	2.4	686	15	US-10-184-634-524	Sequence 524, App
26	36.6	2.3	744	13	US-10-027-632-147012	Sequence 147012,
27	36.6	2.3	744	16	US-10-027-632-147012	Sequence 147012,
c 28	36.2	2.3	2236	16	US-10-377-079-11	Sequence 11, Appl
c 29	36.2	2.3	2488	16	US-10-377-079-9	Sequence 9, Appli
c 30	36	2.3	456	13	US-10-424-599-37000	Sequence 37000, A
31	35.6	2.3	257	9	US-09-867-701-2637	Sequence 2637, Ap
c 32	35.6	2.3	330	9	US-09-880-107-893	Sequence 893, App
c 33	35.6	2.3	440	9	US-09-867-701-10557	Sequence 10557, A
34	35.6	2.3	777	15	US-10-184-644-348	Sequence 348, App
35	35.6	2.3	777	15	US-10-184-634-348	Sequence 348, App
c 36	35.6	2.3	35143	13	US-10-087-192-1894	Sequence 1894, Ap
c 37	35.6	2.3	122186	10	US-09-563-728A-36	Sequence 36, Appl
38	35.4	2.3	1049	13	US-10-142-426-358	Sequence 358, App
39	35.4	2.3	1049	15	US-10-123-155-358	Sequence 358, App
40	35.4	2.3	1049	15	US-10-146-731-358	Sequence 358, App
41	35.4	2.3	1049	15	US-10-140-472-358	Sequence 358, App
42	35.4	2.3	1049	15	US-10-141-761-358	Sequence 358, App
43	35.4	2.3	1049	15	US-10-142-885-358	Sequence 358, App
44	35.4	2.3	1049	15	US-10-158-790-358	Sequence 358, App
45	35.4	2.3	1049	16	US-10-137-871-358	Sequence 358, App
46	35.4	2.3	1049	16	US-10-140-923-358	Sequence 358, App
47	35.4	2.3	1049	16	US-10-141-756-358	Sequence 358, App
48	35.4	2.3	1049	16	US-10-141-759-358	Sequence 358, App
49	35.4	2.3	1049	16	US-10-140-805-358	Sequence 358, App
50	35.4	2.3	1049	16	US-10-140-864-358	Sequence 358, App

ALIGNMENTS

RESULT 1

US-09-989-981A-9

; Sequence 9, Application US/09989981A
 ; Publication No. US20030049730A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hobbs, Helen H.
 ; APPLICANT: Shan, Bei
 ; APPLICANT: Barnes, Robert
 ; APPLICANT: Tian, Hui
 ; APPLICANT: Tularik Inc.
 ; APPLICANT: Board of Regents, The University of Texas System
 ; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
 ; FILE REFERENCE: 018781-007320US
 ; CURRENT APPLICATION NUMBER: US/09/989,981A
 ; CURRENT FILING DATE: 2002-07-23
 ; PRIOR APPLICATION NUMBER: US 60/252,235
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/253,645
 ; PRIOR FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 9
 ; LENGTH: 6043
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: ABCG8 exon 2 (reverse strand) through ABCG5 exon 2
 ; OTHER INFORMATION: (forward strand)
 US-09-989-981A-9

Query Match 99.9%; Score 1568; DB 10; Length 6043;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTCTCAGC	60
Db	3436	CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTCTCAGC	3495
Qy	61	CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCTGCTGTTGGGCTCTCTCTGTC	120
Db	3496	CATGACCAGTGCTGTTTGTGCCCTTTGTGTGGCCTCCCTGCTGTTGGGCTCTCTCTGTC	3555
Qy	121	TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT	180
Db	3556	TTTGCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATT	3615
Qy	181	CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG	240
Db	3616	CCTYTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGG	3675
Qy	241	ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTAT	300
Db	3676	ACACTCTGGCTAAAGGTACATCAGATAATGGCATCGTTGGCCAAATTGGTGAAGTGTAT	3735
Qy	301	CTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT	360
Db	3736	CTCACGAGGATTCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTT	3795

Qy	361	AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCATTTGCTTCCTGCTAG	420
Db	3796	AAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCATTTGCTTCCTGCTAG	3855
Qy	421	CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG	480
Db	3856	CCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAG	3915
Qy	481	GGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAG	540
Db	3916	GGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAG	3975
Qy	541	GTGTCCTGCATGTGTCTTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCT	600
Db	3976	GTGTCCTGCATGTGTCTTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCT	4035
Qy	601	CTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGC	660
Db	4036	CTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGC	4095
Qy	661	AGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTG	720
Db	4096	AGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTG	4155
Qy	721	CGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCAC	780
Db	4156	CGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCAC	4215
Qy	781	TGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGG	840
Db	4216	TGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGG	4275
Qy	841	CTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGG	900
Db	4276	CTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGAAGGAATGCAGG	4335
Qy	901	GTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAA	960
Db	4336	GTTCACTTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAA	4395
Qy	961	TCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACCTTGTCAATTATACCTCC	1020
Db	4396	TCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACCTTGTCAATTATACCTCC	4455
Qy	1021	AAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACT	1080
Db	4456	AAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACT	4515
Qy	1081	GGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTG	1140
Db	4516	GGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTG	4575
Qy	1141	CCCTTGTGGGATTTCTTTAAAGCAACCGTGTGGGCCTTGGTGGAACATCAAATCATGC	1200
Db	4576	CCCTTGTGGGATTTCTTTAAAGCAACCGTGTGGGCCTTGGTGGAACATCAAATCATGC	4635

Qy 1201 CAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAG 1260
 |||
 Db 4636 CAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAG 4695
 Qy 1261 ATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGSCSGGGGCTCCTGTACTTC 1320
 |||
 Db 4696 ATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGSCSGGGGCTCCTGTACTTC 4755
 Qy 1321 TAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTA 1380
 |||
 Db 4756 TAAGGCAGGCTCTGGGAGGCTTTGGCTCYGTCTAAGCACAAATGTTTAAGAAGTRAGTTTA 4815
 Qy 1381 AGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATCTGGTGACTTGTCTGGCT 1440
 |||
 Db 4816 AGTTGTAGAGAGGCAGCCATGCATTTGGCATTGAATACAATCTGGTGACTTGTCTGGCT 4875
 Qy 1441 GCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAG 1500
 |||
 Db 4876 GCCAATAGAACCTAGTACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAG 4935
 Qy 1501 TCCTGCCTAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCA 1560
 |||
 Db 4936 TCCTGCCTAACACGTAAGTGCCTTCTTTGCTTGTTTGATTGACTGTGATGCTAGAGAGCA 4995
 Qy 1561 AACCCAGAGC 1570
 |||
 Db 4996 AACCCAGAGC 5005

RESULT 2

US-09-989-981A-10

; Sequence 10, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

; FILE REFERENCE: 018781-007320US

; CURRENT APPLICATION NUMBER: US/09/989,981A

; CURRENT FILING DATE: 2002-07-23

; PRIOR APPLICATION NUMBER: US 60/252,235

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/253,645

; PRIOR FILING DATE: 2000-11-28

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 10

; LENGTH: 359

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: sequence between ABCG5 and ABCG8 containing

; OTHER INFORMATION: control sequences (bidirectional promoter)

US-09-989-981A-10

Query Match 22.8%; Score 358.6; DB 10; Length 359;
Best Local Similarity 100.0%; Pred. No. 1e-110;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      64  GACCAGTGCTGTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTCTTT 123
          |||||||
Db      1   GACCAGTGCTGTTGTGCCCTTTGTGTGGCCTCCCCTGCTGTTGGGCTCTCTCTGTCTTT 60

Qy     124  GCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCT 183
          |||||||
Db      61  GCTCCTTAGAGCTGGGGCACCTGAGCCCTCCTCTGTGCCAGCCTTTCTCCCAGCATTCCT 120

Qy     184  YTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACA 243
          |||||||
Db     121  YTCTGGCAAACACTTCCTATAAACACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACA 180

Qy     244  CTCTGGCTAAAGGTACATCAGATAAATGGCATCGTTGGCCAAATTGGTGAACGTGTATCTC 303
          |||||||
Db     181  CTCTGGCTAAAGGTACATCAGATAAATGGCATCGTTGGCCAAATTGGTGAACGTGTATCTC 240

Qy     304  ACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAA 363
          |||||||
Db     241  ACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAA 300

Qy     364  GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAGCC 422
          |||||||
Db     301  GCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTTGCAATTTGCTTCCTGCTAGCC 359
```

RESULT 3

US-10-104-047-825

; Sequence 825, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 825
; LENGTH: 2512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-825

Query Match 13.7%; Score 215; DB 16; Length 2512;
Best Local Similarity 54.5%; Pred. No. 2.2e-61;
Matches 576; Conservative 0; Mismatches 450; Indels 31; Gaps 6;

```
Qy      237  AAGGACACTCTGGCTAAAGGTACATCAGATAAATGGCATCGTTGGCCAAATTGGTGAACGT 296
          ||||| | ||||||| || ||||| | |||
```

Db 1 AAGGACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAGTCCCATTCCTG 60
 Qy 297 TTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTC 356
 | | | | | | | | | | | | | | | | | | | | | |
 Db 61 CTGTCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCC 120
 Qy 357 AGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCACCTGCATTTGCTTCCTG 416
 | | | | | | | | | | | | | | | | | | | | | |
 Db 121 AACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTG 180
 Qy 417 CTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAAC 476
 | | | | | | | | | | | | | | | | | | | | | |
 Db 181 TTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAAC 240
 Qy 477 AGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGC 536
 | | | | | | | | | | | | | | | | | | | | | |
 Db 241 AGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGC 297
 Qy 537 TTAGGTGTCTGCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTA 596
 | | | | | | | | | | | | | | | | | | | | | |
 Db 298 CTGGGCATCCTCCATGCCTCCTACAGCGTCAGGTAAGGCAGAGCCC-----TTGCTG 349
 Qy 597 GGCTCTCTGATTGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTC 656
 | | | | | | | | | | | | | | | | | | | | | |
 Db 350 CTGCTGCTCCCCCAGGAGTGCGGGGCCCGGCGCTCACCCCTCTGCTGCCTTTCTTCACTC 409
 Qy 657 CAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCT 716
 | | | | | | | | | | | | | | | | | | | | | |
 Db 410 TTTAAGTGCCAGTCTGGGCACCTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGGAG 469
 Qy 717 CCTGCGCAGGGCAGAGCCTGGACATTAAACATGCCCTGCCTGAAGCCGCTTGCTGCTTC 776
 | | | | | | | | | | | | | | | | | | | | | |
 Db 470 AGGGAGAAGGGCTGTGCTGGGAAACATGGAGCGACAGTGAATGGCCCTCCCCCTGCCCA 529
 Qy 777 TCACTGATTTCTGCTCTCCCCTTCCTTGACTC-GCCCACCACCTGTCCTGTGTAGATGGA 835
 | | | | | | | | | | | | | | | | | | | | | |
 Db 530 GGGAAGGGCCTGGGCATAAACAAAGTGGCAGCAGTGCCCTGCCAACCAGTGTCTACGGC 589
 Qy 836 GAAGGCTC-----GGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTGCTGA 889
 | | | | | | | | | | | | | | | | | | | | | |
 Db 590 CTGCCCTCTGTGGATGGGAATGGGGGTACTGCGAATGCAAGGAGTCTTGAAACCTGGTGA 649
 Qy 890 AGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGTGTGCAGGTGTACCATCTCCCAGTCAG 949
 | | | | | | | | | | | | | | | | | | | | | |
 Db 650 AAGAATGCAGGG-----ACAGCCACCTCGCAGCCAAACGGACAGGACATTTCAGAGCAAC 703
 Qy 950 AGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGGCCAACTTGTC 1009
 | | | | | | | | | | | | | | | | | | | | | |
 Db 704 TCCAGCACAGGCCCCCTCCCTACGTGGCAGACAGCCTCAGTCGCTATCTGCCAGGTTCT- 762
 Qy 1010 ATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAACCTGTC 1069
 | | | | | | | | | | | | | | | | | | | | | |
 Db 763 -----ACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCCTGTCCGGAGACTACTG 816
 Qy 1070 CAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCAGGACCCTA 1129
 | | | | | | | | | | | | | | | | | | | | | |
 Db 817 GGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGCCCCTTCAGGGCCCCA 876

Qy	378	GGACAGGCCACTAGAAAATTCACTTGCAATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCC	437
Db	2	GGACAGGCCACTAGAAAATTCACTTGCAATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCC	61
Qy	438	TTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTG	497
Db	62	TTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTG	121


```

Qy      498 GAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCC 557
          |||
Db      122 GAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCC 181

Qy      558 TACAGCGTCAGGTAAGGGGAC 578
          |||
Db      182 TACAGCGTCAGCAACCGTGTC 202

```

RESULT 5

US-09-989-981A-1

```

; Sequence 1, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1959)
; OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-1

```

```

Query Match          9.3%; Score 146.4; DB 10; Length 1959;
Best Local Similarity 96.2%; Pred. No. 4.2e-38;
Matches 150; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy      423 ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 482
          |||
Db      1   ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 60

Qy      483 TCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 542
          |||
Db      61 TCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120

Qy      543 GTCCTGCATGTGTCCTACAGCGTCAGGTAAGGGGAC 578
          |||
Db      121 GTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTC 156

```

RESULT 6

US-09-837-992-7

; Sequence 7, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 249
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 1 of hSSG
US-09-837-992-7

Query Match 6.5%; Score 101.6; DB 9; Length 249;
Best Local Similarity 68.4%; Pred. No. 2.2e-23;
Matches 156; Conservative 0; Mismatches 69; Indels 3; Gaps 1;

Qy 341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
|| ||| ||||| || | |||| | |||| | || | ||||| ||||| |
Db 25 CTGCCACGGGCTCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
Qy 401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
||||| |||| | ||||| |||| | | | ||| || | || |
Db 85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT 144
Qy 461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGTCACGGGCAC 520
|| | || | ||||| || | ||||| ||||| | | | ||| || |
Db 145 GGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
Qy 521 AGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAG 568
||| || ||||| | || |||| |||| ||||| |||||
Db 205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAG 249

RESULT 7

US-09-837-992-4

; Sequence 4, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua

```

; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human sitosterolemia gene (SSG)
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: protein
US-09-837-992-4

```

```

Query Match          6.5%; Score 101.6; DB 9; Length 2340;
Best Local Similarity 67.4%; Pred. No. 8e-23;
Matches 159; Conservative 0; Mismatches 74; Indels 3; Gaps 1;

```

```

Qy      341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
          || ||| ||||| || | |||| | |||| | || ||||| ||||| |
Db      25  CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84

Qy      401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
          ||||| |||| | ||||| ||||| || | | ||| || | |||
Db      85  CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT 144

Qy      461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCAC 520
          || | || | ||||| || | ||||| ||||| | | | ||| | | |
Db      145 GGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204

Qy      521 AGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCCTACAGCGTCAGGTAAGGGG 576
          ||| || ||||| | || |||| ||| ||||| ||||| | | |
Db      205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 257

```

RESULT 8

US-09-989-981A-5

```

; Sequence 5, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System

```

```
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-5
```

```
Query Match          6.5%; Score 101.6; DB 10; Length 2340;
Best Local Similarity 67.4%; Pred. No. 8e-23;
Matches 159; Conservative 0; Mismatches 74; Indels 3; Gaps 1;
```

```
Qy      341 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 400
      || ||| ||||| || | ||| | ||| | ||| ||||| ||||| |
Db      25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84

Qy      401 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 460
      ||||| |||| | ||||| ||| || | || | || | |||
Db      85 CCAGCTTTGCTGCCTGTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCAT 144

Qy      461 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCAC 520
      || | || | ||||| || | ||||| ||| | || | || | |||
Db      145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204

Qy      521 AGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTACAGCGTCAGGTAAGGGG 576
      ||| || | |||| | || ||| ||| ||||| ||||| | | |
Db      205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCG 257
```

RESULT 9

US-09-837-992-8

```
; Sequence 8, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
```

; PRIOR APPLICATION NUMBER: US 60/204,234
 ; PRIOR FILING DATE: 2000-05-15
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 8
 ; LENGTH: 122
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: exon 2 of hSSG
 US-09-837-992-8

Query Match 5.7%; Score 90; DB 9; Length 122;
 Best Local Similarity 83.6%; Pred. No. 1.3e-19;
 Matches 102; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1164 CAACCGTGTCTGGGCCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
 | |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1 CCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGAT 60
 Qy 1224 CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTC 1283
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 61 CCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTC 120
 Qy 1284 AG 1285
 ||
 Db 121 AG 122

RESULT 10

US-09-989-981A-3/c
 ; Sequence 3, Application US/09989981A
 ; Publication No. US20030049730A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hobbs, Helen H.
 ; APPLICANT: Shan, Bei
 ; APPLICANT: Barnes, Robert
 ; APPLICANT: Tian, Hui
 ; APPLICANT: Tularik Inc.
 ; APPLICANT: Board of Regents, The University of Texas System
 ; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
 ; FILE REFERENCE: 018781-007320US
 ; CURRENT APPLICATION NUMBER: US/09/989,981A
 ; CURRENT FILING DATE: 2002-07-23
 ; PRIOR APPLICATION NUMBER: US 60/252,235
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/253,645
 ; PRIOR FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 2019
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2019)

; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-3

Query Match 4.0%; Score 63; DB 10; Length 2019;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 60
|||||
Db 63 CGAAGCATCCTGAAGTACAGTCCCATTCCACAGCTGGGTCTCTTCTTTGGTTTTCTCAGC 4

Qy 61 CAT 63
|||
Db 3 CAT 1

RESULT 11

US-10-142-426-412

; Sequence 412, Application US/10142426
; Publication No. US20040048333A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C224

; CURRENT APPLICATION NUMBER: US/10/142,426

; CURRENT FILING DATE: 2002-05-09

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 412

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-142-426-412

Query Match 2.5%; Score 39.2; DB 13; Length 1184;
Best Local Similarity 9.3%; Pred. No. 0.092;
Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

Qy 449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC 508
::|| ::| :| | : : : : : :

Db 320 RRAGQSVSLCCKATGKPRPDKYFWYHNDTLLDPSLYKHESKLVLRKLQQHQAGEYFCKAQ 379
 Qy 509 GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTACAGCGTCAG 568
 :: |:::| : :: | : : : | || :: :|
 Db 380 SDAGAVKSKVAQLIVTASDETPCNVPESYLIRLPHDCFQATNSFYDVGRCPVKTCAG 439
 Qy 569 GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
 : | : : | : : : :| : : : | | : : :
 Db 440 QQDNGIRCRDAVQNCCGISKTEEREIQCSGYTLPTKVAKECSCQRCCTETRSIVRGRVSAA 499
 Qy 625 -TGGGTGGGCCTGTGGGCTTTGGGTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
 | : |:: : : || : | : : : | : : : :| : :
 Db 500 DNGEPMRFHVMGNSRVSMGTGYKGTFTLHVPQDTERLVLTfVDRLQKFVNTTKVLPFNK 559
 Qy 684 CTGTAACAACAGTGAGTCGTTCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA 743
 : : : : : : : : : : : : : : |
 Db 560 KGSVAFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK 619
 Qy 744 AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTT 803
 : |::| : | : : : :| : : : : : : : :
 Db 620 VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSVDfRDEVTSEPLNAGKVK 679
 Qy 804 GACTCGCCCACCACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG 863
 : : : : : : :| : : : | : : : : : : : :
 Db 680 VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER 739
 Qy 864 GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGT 923
 : : : : : : | : : : : : : | : : : : : : :
 Db 740 RLFNLDVPESRRCFVKVRAYRSERFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799
 Qy 924 GTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
 || | : | |:: : : : | : : : : : : : :
 Db 800 ITGPNGACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLN 859
 Qy 984 GCTCCTTGGGTGGTGGCCAACTTGTCTATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
 : | : : : : : :| : : : | : : : :| : : :
 Db 860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919
 Qy 1044 GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
 : : : : : : : : : | : : : | : : : || :
 Db 920 DRYDYNTPFNEDDPMSWTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979
 Qy 1104 GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTTAAAG 1163
 :| : : : :| : : : : : : : : :| : : : :| :
 Db 980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLKVIPQGSRRASVN 1039
 Qy 1164 CAACCGTGTGGGCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
 : : : : : : : | : : : : : : : :| : | : |
 Db 1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099
 Qy 1224 C 1224
 :
 Db 1100 S 1100

```

; Sequence 412, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 412
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-412

```

```

Query Match          2.5%; Score 39.2; DB 15; Length 1184;
Best Local Similarity 9.3%; Pred. No. 0.092;
Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

```

```

Qy      449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC 508
        ::|| ::| :| | : : : : : : : : : : : : : : : :
Db      320 RRAGQSVSLCCKATGKPRDPKYFWYHNDTLLDPSLYKHESKLVLRLKLQQHQAGEYFCKAQ 379

Qy      509 GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCCTACAGCGTCAG 568
        :: |:::| : : : | : : : | | : : : : : : : :
Db      380 SDAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDCFQ NATNSFY YDVGRCPVKTCAG 439

Qy      569 GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
        : | : : | : : : :| : : : :| | : : : : : :
Db      440 QQDNGIRCRDAVQNC CGISKTEEREIQCSGYTLPTKVAKESCQRCTETRSIVRGRVSAA 499

Qy      625 -TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
        | : |:: : : || : | : : : | : : : : : : : :
Db      500 DNGEPMRFGHVYMGNSRVSM TGYKGTFTLHVPQDTERLVLTFVDRLQKFVNTTKVLPFNK 559

Qy      684 CTGTAACAACAGTGAGTCGTTCTCTCTCTCTCTCTGCGCAGGGCAGAGCCTGGACATTA 743
        : : : : : : : : : : : : : : : |
Db      560 KGS AVFHEIKMLRRKEPIT LEAMETNI IPLGEVVGEDPMAELEIPSR SFYRQNGEPYIGK 619

```


Qy 744 AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTT 803
: |::| :| | : :| : :|: : :|: :
Db 620 VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSVDFRDEVTSEPLNAGKVK 679

Qy 804 GACTCGCCCACCACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG 863
:: : : : :|:: : | : | : : : : : : : : :
Db 680 VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER 739

Qy 864 GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTTCAAGAAGAAAGCAGT 923
: : : : :|: : : |:: : : | : :|: : : :
Db 740 RLFNLDVPESRRCFVKVRAYRSEFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799

Qy 924 GTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
|| | : | |:: : : :|: : | : : : : : : :
Db 800 ITGPNGACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLNLY 859

Qy 984 GCTCCTTGGGTGGTGGCCAACTTGTCAATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
: | : : : : : :| : : : : : : :| : : : :
Db 860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919

Qy 1044 GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
:::: : : : : : | : : : | : : : || :
Db 920 DRYDYNTVPFNEDDPMSWTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979

Qy 1104 GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCCCTTTAAAG 1163
:|: : : : :| : : : : : : : :| : : : : | :
Db 980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLVKVIPQGSCRRASVN 1039

Qy 1164 CAACCGTGTCTGGGCCTTGGTGGAACATCAAATCATGCCAGCAGAAAGTGGGACAGGCAAAT 1223
: : : : : : : | : : : :| : : : :| : :| :| :|
Db 1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099

Qy 1224 C 1224
:
Db 1100 S 1100

RESULT 13

US-10-146-731-412

; Sequence 412, Application US/10146731

; Publication No. US20030129692A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

```

; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C323
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 412
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-146-731-412

```

```

Query Match          2.5%; Score 39.2; DB 15; Length 1184;
Best Local Similarity 9.3%; Pred. No. 0.092;
Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

```

```

Qy      449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC 508
      ::|| ::| :| | : : : : : : : : : : : : : : : :
Db      320 RRAGQSVSLCCKATGKPRPKYFWYHNDTLLDPSLYKHESKLVLRKLQQHQAGEYFCKAQ 379

Qy      509 GGTACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAG 568
      :: |:::| : : : | : : : : | || :: :|||
Db      380 SDAGAVKSKVAQLIVTASDETPCNVPESYLIRLPHDCFQATNSFYVDVGRCPVKTCAG 439

Qy      569 GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
      : | : : | : : : :| : :| :| :| :| :| :| :|
Db      440 QQDNGIRCRDAVQNCCKISKEEREIQCSGYTLPTKVAKECSCQRCCTETRSIVRGRVSAA 499

Qy      625 -TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
      | : |:: : : || : | | : : : | : : : : :| : :
Db      500 DNGEPMRFGHVYMGNSRVSMTRYKGTFTLHVPQDTERLVLTFFVDRLQKFVNTTKVLPFNK 559

Qy      684 CTGTAACAACAGTGAGTCGTTTCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA 743
      : : : : : : : : : : : : : : : |
Db      560 KGSADFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK 619

Qy      744 AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTT 803
      : ||:| :| | : : : :| : :| : : :| : :
Db      620 VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSDFRDEVTSEPLNAGKVK 679

Qy      804 GACTCGCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG 863
      : : : : : :| : : : | : : : : : : : : : :
Db      680 VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER 739

Qy      864 GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGT 923
      : : : : :| : : : | : : : | : : : : :
Db      740 RLFNLDVPESRRCFVKVRAYSERFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799

Qy      924 GTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
      || | : | | : : : :| : : : : : : : : :
Db      800 ITGPNACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLNLY 859

Qy      984 GCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
      : | : : : : : :| :| : : : :| : :

```

Db	860	RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG	919
QY	1044	GGCTAAACACAGATTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA	1103
		:::: : :: : : : :: : : :	
Db	920	DRYDYNTVPFNEDDPMSWTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR	979
QY	1104	GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAG	1163
		: : : : : : : : : : : : : : : :	
Db	980	TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLVKVIPQGSCRRASVN	1039
QY	1164	CAACCGTGTGCGGCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT	1223
		: : : : : : : : : : : : : : : : :	
Db	1040	PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT	1099
QY	1224	C 1224	
		:	
Db	1100	S 1100	

US-10-140-472-412

: Publication No. US20030138888A1

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

: APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

: APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Zhang, Zemin

: TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C168

: CURRENT APPLICATION NUMBER: US/10/140,472

; CURRENT FILING DATE: 2002-05-06

: NUMBER OF SEO ID NOS: 550

: SEO ID NO 412

```

,   LENGTH: 1104
:   TYPE: PRT

```

: ORGANISM: Homo Sapien

Query Match 2.5%; Score 39.2; DB 15; Length 1184;

Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

QY	449	AGAGGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC	508
Db	320	RRAGQSVSLCCKATGKPRPDKYFWYHNDTLLDPSLYKHESKLVLRKLQHQAGEYFCKAQ	379
QY	509	GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTACAGCGTCAG	568
Db	380	SDAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDCFQATNSFYDVGRCVPKTCAG	439
QY	569	GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG----	624
Db	440	QQDNGIRCRDAVQNC CGISKTEEREIQCSGYTLPTKVAKECSCQRCTETRSIVRGRVSAA	499
QY	625	-TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT	683
Db	500	DNGEPMRFHGVYMGNSRVSM TGYKGTFTLHVPQDTERLVLTFVDRLQKFVNTTKVLPFNK	559
QY	684	CTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA	743
Db	560	KGSAVFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK	619
QY	744	AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTT	803
Db	620	VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSVD FRDEV TSEPLNAGKVK	679
QY	804	GACTCGCCCAACACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG	863
Db	680	VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER	739
QY	864	GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTTCAAGAAGAAAGCAGT	923
Db	740	RLFNLDVPESRRCFVKVRAYR SERFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV	799
QY	924	GTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT	983
Db	800	ITGPN GACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNALGVPQPYLNKLN Y	859
QY	984	GCTCCTTGGGTGGTGGCCA ACTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAA	1043
Db	860	RRTDHEDPRVKKTAFQISM AKPRPN SAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG	919
QY	1044	GGCTAAAAACAGATTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA	1103
Db	920	DRYDNTVPFNEDDPMSWTE DYLA WWP KPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR	979
QY	1104	GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTG TGGGATTCCTTTAAAG	1163
Db	980	TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTL VKVIPQGSCRRASVN	1039
QY	1164	CAACCGTGTCGGGCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAA T	1223
Db	1040	PMLHEYLVNHLPLAVNNDTSEYTM LAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT	1099
QY	1224	C 1224	
Db	1100	S 1100	

RESULT 15

US-10-141-761-412

; Sequence 412, Application US/10141761

; Publication No. US20030148432A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C198

; CURRENT APPLICATION NUMBER: US/10/141,761

; CURRENT FILING DATE: 2002-05-08

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 412

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-141-761-412

Query Match 2.5%; Score 39.2; DB 15; Length 1184;

Best Local Similarity 9.3%; Pred. No. 0.092;

Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

```

Qy      449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC 508
          ::|| ::| :| |:      : ::      : : :: : :
Db      320 RRAGQSVSLCCKATGKPRPDYFWYHNDTLLDPSLYKHESKLVLRLKQQHQAGEYFCKAQ 379

Qy      509 GGTACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAG 568
          :: |:::| : : : | : : : | || :: ::|||
Db      380 SDAGAVKSKVAQLIVTASDETPCNVPESYLIRLPHDCFQATNSFYVDVGRCPVKTCAG 439

Qy      569 GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
          : | : : | : : : :| :| :| :| :| :| :|
Db      440 QQDNGIRCRDAVQNCCGISKTEEREIQCSGYTLPTKVAKECSCQRCCTETRSIVRGRVSAA 499

Qy      625 -TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
          | :|:: : : || :| :| : :| :| : :| :| :
Db      500 DNGEPMRFHGYMGNSRVSMYGYKGTFTLHVPQDTERLVLTFTVDRLLQKFVNTTKVLPFNK 559

Qy      684 CTGTAACAACAGTGAGTCGTTTCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA 743

```

```

      : : : : : : : : : :
Db      560 KGS AVFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK 619

Qy      744 AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTT 803
      : |::| : | : : : | : : | : : : :
Db      620 VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSVDFRDEVTSEPLNAGKVK 679

Qy      804 GACTCGCCCACCACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG 863
      :: : : : : |:: : | : | : : : : : : :
Db      680 VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER 739

Qy      864 GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGT 923
      : : : : : | : : : | : : : | : : : :
Db      740 RLFNLDVPESRRCFVKVRAYRSEFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799

Qy      924 GTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
      || | : | |:: : : : | : : : : : : :
Db      800 ITGPNACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLNLY 859

Qy      984 GCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
      : | : : : : : : | : : : : : : :
Db      860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919

Qy      1044 GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
      : : : : : : : : : : | : : : : | : : : :
Db      920 DRYDYNTVPFNEDDPMSWTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979

Qy      1104 GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAG 1163
      : | : : : : | : : : : : : : : : | : : : :
Db      980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLVKVIPQGSRRASVN 1039

Qy      1164 CAACCGTGTGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
      : : : : : : : | : : : : : : : : : | : : : :
Db      1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099

Qy      1224 C 1224
      :
Db      1100 S 1100

```

RESULT 16

US-10-142-885-412

; Sequence 412, Application US/10142885

; Publication No. US20030157604A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria

Qy 984 GCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
 : | : : : : : : : | : | : : : : : : :
 Db 860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919
 Qy 1044 GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
 : : : : : : : : : : : : : : : : : : :
 Db 920 DRYDYNTVPFNEDDPMSWTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979
 Qy 1104 GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTTAAAG 1163
 : | : : : : : : : : : : : : : : : : : :
 Db 980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLKVVIPQGSCRRASVN 1039
 Qy 1164 CAACCGTGTCTGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
 : : : : : : : : : : : : : : : : : : :
 Db 1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099
 Qy 1224 C 1224
 :
 Db 1100 S 1100

RESULT 17

US-10-158-790-412

; Sequence 412, Application US/10158790

; Publication No. US20030180879A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C448

; CURRENT APPLICATION NUMBER: US/10/158,790

; CURRENT FILING DATE: 2002-05-30

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 412

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-158-790-412

Query Match 2.5%; Score 39.2; DB 15; Length 1184;
 Best Local Similarity 9.3%; Pred. No. 0.092;
 Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

Qy	449	AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC	508
		:: :: : : : :: : : : : :	
Db	320	RRAGQSVSLCCKATGKPRPDKYFWYHNDTLDDPSLYKHESKLVLRKLQQHQAGEYFCKAQ	379
Qy	509	GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTACAGCGTCAG	568
		:: ::: : : : : : : :: :	
Db	380	SDAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDCFQATNSFYVDVGRCPVKTCAG	439
Qy	569	GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG----	624
		: : : : : : : : : : : :	
Db	440	QQDNGIRCRDAVQNCGISKTEEREIQCSGYTLPTKVAKECSCQRCCTETRSIVRGRVSAA	499
Qy	625	-TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT	683
		: :: : : : : : : : : : : : :	
Db	500	DNGEPMRFGHVYMGNSRVSMTRYKGTFTLHVPQDTERLVLTFFVDRLQKFVNTTKVLPFNK	559
Qy	684	CTGTAACAACAGTGAGTCGTTCTCTCTCTCTCTCGCGCAGGGCAGAGCCTGGACATTA	743
		: : : : : : : : : : : :	
Db	560	KGSAVFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK	619
Qy	744	AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCTCT	803
		: :: : : : : : : : : : : :	
Db	620	VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRITYGMFSVDFRDEVTSEPLNAGKVK	679
Qy	804	GACTCGCCACCACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG	863
		:: : : : : : : : : : : : : : : :	
Db	680	VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER	739
Qy	864	GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGT	923
		: : : : : : : : : : : : : : : :	
Db	740	RLFNLDVPESRRCFVKVRAYRSEFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV	799
Qy	924	GTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT	983
		: : : : : : : : : : : :	
Db	800	ITGPNGACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLN	859
Qy	984	GCTCCTTGGGTGGTGGCCAACCTGTCTATTATACCTCCAAGGACAACAGAGTGGTACATAA	1043
		: : : : : : : : : : : : : :	
Db	860	RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG	919
Qy	1044	GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA	1103
		::: : : : : : : : : : : : : : :	
Db	920	DRYDINTVPFNEDDPMSWTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR	979
Qy	1104	GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAG	1163
		: : : : : : : : : : : : : : : :	
Db	980	TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLKVPIPGSCRRASVN	1039
Qy	1164	CAACCGTGTGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT	1223
		: : : : : : : : : : : : : : : :	
Db	1040	PMLHEYLVNHLPLAVNNDTSEYTMPLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT	1099
Qy	1224	C	1224

Db :
 1100 S 1100

RESULT 18

US-10-137-871-412

; Sequence 412, Application US/10137871

; Publication No. US20030207350A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C153

; CURRENT APPLICATION NUMBER: US/10/137,871

; CURRENT FILING DATE: 2002-05-03

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 412

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-137-871-412

Query Match 2.5%; Score 39.2; DB 16; Length 1184;

Best Local Similarity 9.3%; Pred. No. 0.092;

Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

```
Qy      449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTC 508
          ::|| ::| :| | : : : : : : : : : : : : : : : :
Db      320 RRAGQSVSLCCKATGKPRPDKYFWYHNDTLLDPSLYKHESKLVLRLQQHQAQGEYFCKAQ 379

Qy      509 GGTACAGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAG 568
          :: |:::| : : : | : : : : | | :: :|::|
Db      380 SDAGAVKSKVAQLIVTASDETPCNVPESYLIRLPHDCFQATNSFYDVGRCVPKTCAG 439

Qy      569 GTAAGGGGACCTCCACAGCAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
          : | : : : | : : : :| : :| :| :| :| :|
Db      440 QQDNGIRCRDAVQNCCGISKTEEREIQCSGYTLPTKVAKECSCQRCCTETRIVRGRVSAA 499

Qy      625 -TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
          | :|:: : : || :| | : : :| : : : :| : :
```



```

; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 412
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-923-412

```

```

Query Match          2.5%; Score 39.2; DB 16; Length 1184;
Best Local Similarity 9.3%; Pred. No. 0.092;
Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

```

```

Qy      449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTC 508
          ::|| ::| :| |:      : ::      : : :: : :      :
Db      320 RRAGQSVSLCKATGKPRDPKYFWYHNDTLLDPSLYKHESKLVLRKLQQHQAGEYFCKAQ 379

Qy      509 GGTACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTTACAGCGTCAG 568
          :: |:::| : : : | : : : : | || :: :| |||
Db      380 SDAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDCFQATNSFYDVGRCVPVKTCAG 439

Qy      569 GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
          : | : :      | : : : :| :| ::| : :| | :| :| :
Db      440 QQDNGIRCRDAVQNCCGISKTEEREIQCSGYTLPTKVAKECSCQRCCTETRSIVRGRVSAA 499

Qy      625 -TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
          | : |:: : : : || : | | : :      : | :: : : | : :
Db      500 DNGEPMRFGHVYMGNSRVSMGTGYKGTFTLHVPQDTERLVLTFVDRLQKFVNTTKVLPFNK 559

Qy      684 CTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA 743
          : : : : : : : : : : : : : : : |
Db      560 KGSAVFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK 619

Qy      744 AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTT 803
          : |::| :| | : : : | : :| : : :| : :
Db      620 VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRITYGMFSVDFRDEVTSEPLNAGKVK 679

Qy      804 GACTCGCCACACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG 863
          :: : : : :|:: : | : | : : : : : : : : :
Db      680 VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER 739

Qy      864 GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGT 923
          : : : : :| : : : | : : : | : : : :
Db      740 RLFNLDVPESRRCFVKVRAYRSEFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799

```

Qy 924 GTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
 || | : | |:: : :: | : : : :
 Db 800 ITGPNACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPPYLNKLN 859
 Qy 984 GCTCCTTGGGTGGTGGCCAACTTGTCAATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
 : | : : :::: : : | : : : || :
 Db 860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919
 Qy 1044 GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
 :::: : : : : | : :: | : : : || :
 Db 920 DRYDYNTVPFNEDDPMSTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979
 Qy 1104 GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTTCCTTTAAAG 1163
 : | : : : : | : : : : : : : : : : : | :
 Db 980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLVKVIPQGSRRASVN 1039
 Qy 1164 CAACCGTGTGCGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAAGTGGGACAGGCAAAT 1223
 : : : : : : : | : : : : : : : : : | : | : |
 Db 1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099
 Qy 1224 C 1224
 :
 Db 1100 S 1100

RESULT 20

US-10-141-756-412

; Sequence 412, Application US/10141756

; Publication No. US20030207359A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C200

; CURRENT APPLICATION NUMBER: US/10/141,756

; CURRENT FILING DATE: 2002-05-08

; Prior Apploication removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 412

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: Homo Sapien
US-10-141-756-412

Query Match 2.5%; Score 39.2; DB 16; Length 1184;
Best Local Similarity 9.3%; Pred. No. 0.092;
Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

```
Qy      449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC 508
          ::|| :::: | :| | :      : ::::      : : :::: : :      :
Db      320 RRAGQSVSLCKATGKPRPDKYFWYHNDTLLDPSLYKHESKLVLRKLQQHQAGEYFCKAQ 379

Qy      509 GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTTACAGCGTCAG 568
          :: | ::::| : : : | : : : : : | | : : : : : |||
Db      380 SDAGAVKSKVAQLIVTASDETPCNVPESYLIRLPHDCFQATNSFYVDVGRCPVKTCAG 439

Qy      569 GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
          : | : : : | : : : : : : | : : : : | | : : : : :
Db      440 QQDNGIRCRDAVQNCCGISKTEEREIQCSGYTLPTKVAKESCQRCTETRSIVRGRVSAA 499

Qy      625 -TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
          | : | :::: :::: || : | | :: : | :::: : : : | :: :
Db      500 DNGEPMRFHGYMGNSRVSMGTGYKGTFTLHVPQDTERLVLTFFVDRLQKEVNTTKVLPFNK 559

Qy      684 CTGTAACAACAGTGAGTCGTTCTCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA 743
          : : : : : : : : : : : : : : : : : |
Db      560 KGSVAFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK 619

Qy      744 AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTT 803
          : | ::| : | | : : : | : : : : : : : : : : :
Db      620 VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSVDFRDEVTSEPLNAGKVK 679

Qy      804 GACTCGCCACCACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG 863
          :: : : : : : : : : : : : : : : : : : : : : :
Db      680 VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKRDRFTLVGNLEIRER 739

Qy      864 GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTCAAGAAGAAAGCAGT 923
          : : : : : : : : : : : : : : : : : : : : :
Db      740 RLFNLDVPESRRCFVKVRAYSERFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799

Qy      924 GTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
          || | : | | :: : : : | : : : : : : : : : : :
Db      800 ITGPNGACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLN 859

Qy      984 GCTCCTTGGGTGGTGGCCAACTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
          : | : : :::: : : | : : : : : : : : : : :
Db      860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919

Qy      1044 GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
          :::: : : : : : : : : : : : : : : : : : : :
Db      920 DRYDYNTPVFNEDDPMSTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979

Qy      1104 GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAG 1163
          : | : : : : : : : : : : : : : : : : : : :
Db      980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLKVKVIPOGSCRRASVN 1039

Qy      1164 CAACCGTGTGGGCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
          : : : : : : : : : : : : : : : : : : : |
```

Db 1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099
Qy 1224 C 1224
:
Db 1100 S 1100

RESULT 21

US-10-141-759-412

; Sequence 412, Application US/10141759

; Publication No. US20030207361A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C197

; CURRENT APPLICATION NUMBER: US/10/141,759

; CURRENT FILING DATE: 2002-05-08

; Prior Apploication removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 412

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-141-759-412

Query Match 2.5%; Score 39.2; DB 16; Length 1184;

Best Local Similarity 9.3%; Pred. No. 0.092;

Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

Qy 449 AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC 508
::|| ::| :| |: : :: : : : : :

Db 320 RRAGQSVSLCCKATGKPRPDYFWYHNDTLLDPSLYKHESKLVLRKLQQHQAGEYFCKAQ 379

Qy 509 GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTACAGCGTCAG 568
:: |:::| : : : | : : : | || :: :|

Db 380 SDAGAVKSKVAQLIVTASDETPCNVPESYLIRLPHDCFQATNSFYDVGRCVPKTCAG 439

Qy 569 GTAAGGGGACCTCCACAGCAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG---- 624
: | : : | : : :| :| :| :| :|

Db 440 QQDNGIRCDVQNCGGISKTEEREIQCSGYTLPTKVAKECSCQRCCTETRSIVRGRVSAA 499

Qy 625 -TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT 683
 | : |:: : : | | : | : : : | : : : : | : : :
 Db 500 DNGEPMRFGHVYMGNSRVSM TGYKGTFTLHVPQDTERLVLT FVDRLQKFVNTTKVLPFNK 559

Qy 684 CTGTAACAACAGTGAGTCGTTCTCTCTCTCTCTCGCGAGGGCAGAGCCTGGACATTA 743
 : : : : : : : : : : : : : : |
 Db 560 KGS AVFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK 619

Qy 744 AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTT 803
 : |::| : | : : : : | : : : : : : : : : : : :
 Db 620 VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSVDFRDEV TSEPLNAGKVK 679

Qy 804 GACTCGCCACACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG 863
 :
 Db 680 VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER 739

Qy 864 GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACCTTCAAGAAGAAAAGCAGT 923
 :
 Db 740 RLFNLDVPESRRCFVKVRAYRSE RFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799

Qy 924 GTGCAGGTGTACCATCTCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
 || | : | |:: : : : : : : : : : : : : : : : : :
 Db 800 ITGPN GACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLN Y 859

Qy 984 GCTCCTTGGGTGGTGGCCAACCTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
 : | :
 Db 860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919

Qy 1044 GGCTAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
 :
 Db 920 DRYDYN TVPFNEDDPMSWTE DYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979

Qy 1104 GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAG 1163
 : | :
 Db 980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTL VKVIPQGSRRASVN 1039

Qy 1164 CAACCGTGTGGGCCTTGGTGG AACATCAAATCATGCCAGCAGAAAGTGGGACAGGCAAAT 1223
 :
 Db 1040 PMLHEYLVNHLPLAVNNDTSEY TMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099

Qy 1224 C 1224
 :
 Db 1100 S 1100

RESULT 22

US-10-140-805-412

; Sequence 412, Application US/10140805

; Publication No. US20030207417A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3330R1C176
 ; CURRENT APPLICATION NUMBER: US/10/140,805
 ; CURRENT FILING DATE: 2002-05-07
 ; Prior Apploication removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 412
 ; LENGTH: 1184
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-140-805-412

Query Match 2.5%; Score 39.2; DB 16; Length 1184;
 Best Local Similarity 9.3%; Pred. No. 0.092;
 Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

Qy	449	AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC	508
		:: :: : : : : : : : : :	
Db	320	RRAGQSVSLCCKATGKPRPDYFWYHNDTLLDPSLYKHESKLVLRKLQQHQAGEYFCKAQ	379
Qy	509	GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGCGATGTGTCTACAGCGTCAG	568
		:: ::: : : : : : : :: :	
Db	380	SDAGAVKSKVAQLIVTASDETPCNFPVPEYLIRLPHDCFQATNSFYDVGRCVPKTCAG	439
Qy	569	GTAAGGGGACCTCCACAGCAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG----	624
		: : : : : : : : : : :	
Db	440	QQDNGIRCRDAVQNCGISKTEEREIQCSGYTLPTKVAKECSCQRCETRSIVRGRVSAA	499
Qy	625	-TGGGTGGGCCTGTGGGCTTTGGGTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT	683
		: :: : : : : : : : : : : :	
Db	500	DNGEPMRFHVMGNSRVSMGTGYKGTFTLHVPQDTERLVLTFFVDRQLQKFVNTTKVLPFNK	559
Qy	684	CTGTAACAACAGTGAGTCGTTCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA	743
		: : : : : : : : :	
Db	560	KGSAVFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK	619
Qy	744	AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTT	803
		: :: : : : : : : : : : :	
Db	620	VKASVTFLDPRNISTATAAQTDLNFINDGDTFPLRTYGMFSVDFRDEVTSEPLNAGKVK	679
Qy	804	GACTCGCCACACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG	863
		:: : : : : : : : : : : :	
Db	680	VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER	739
Qy	864	GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGT	923

```

      :   : : : : : | : :   :   | : : :   |   :   : | : : :
Db      740 RLFNLDVPESRRCFVKVRAYRSERFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV 799

Qy      924 GTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT 983
      || | : | | : :   : : : | :   | : :   :   : :
Db      800 ITGPNACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLN 859

Qy      984 GCTCCTTGGGTGGTGGCCAACTTGTTCATTATACCTCCAAGGACAACAGAGTGGTACATAA 1043
      : | : : : : : : : |   : |   :   : : | :
Db      860 RRTDHEDPRVKKTAFQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG 919

Qy      1044 GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA 1103
      : : : : : : : : : :   | : : : : |   : : : | :
Db      920 DRYDYNTPFNEDDPMSWTEDYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR 979

Qy      1104 GGGGTCTGGCACCTTCCAGGACCTACTCTGCCTTTGCCCTTGTGGGATTTCTTTAAAG 1163
      : | : : : : : | :   : :   : :   : : : | : : : : | :
Db      980 TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLVKVIPQGSCRRASVN 1039

Qy      1164 CAACCGTGTGGGCCTTGGTGGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
      : : : : : : : : | :   : : : | : : : : : | :   | : : |
Db      1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099

Qy      1224 C 1224
      :
Db      1100 S 1100

```

RESULT 23

US-10-140-864-412

; Sequence 412, Application US/10140864

; Publication No. US20030207419A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C184

; CURRENT APPLICATION NUMBER: US/10/140,864

; CURRENT FILING DATE: 2002-05-07

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

```
; SEQ ID NO 412
;   LENGTH: 1184
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-864-412
```

Query Match 2.5%; Score 39.2; DB 16; Length 1184;
Best Local Similarity 9.3%; Pred. No. 0.092;
Matches 73; Conservative 219; Mismatches 484; Indels 5; Gaps 1;

QY	449	AGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTC	508
Db	320	RRAGQSVSLCCKATGKPRPDKYFWYHNDTLLDPSLYKHESKLVLRKLQQHQAGEYFCKAQ	379
QY	509	GGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTTCATGTGTCTACAGCGTCAG	568
Db	380	SDAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDCFQATNSFYFDVGRCPVKTCAG	439
QY	569	GTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGG----	624
Db	440	QQDNGIRCRDAVQNCCGISKTEEREIQCSGYTLPTKVAKECSCQRCTETRIVRGRVSAA	499
QY	625	-TGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGT	683
Db	500	DNGEPMRFHGHVYMGNSRVSMGTGKGTFTLHVPQDTERLVLTFVDRLQKFVNTTKVLPFNK	559
QY	684	CTGTAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTA	743
Db	560	KGSAVFHEIKMLRRKEPITLEAMETNIIPLGEVVGEDPMAELEIPSRSFYRQNGEPYIGK	619
QY	744	AAACATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTT	803
Db	620	VKASVTFLDPRNISTATAAQTDLNFINDEGDTFPLRTYGMFSVDFRDEVTSEPLNAGKVK	679
QY	804	GACTCGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGG	863
Db	680	VHLDSTQVKMPEHISTVKLWSLNPDTGLWEEEGDFKFENQRRNKREDRTFLVGNLEIRER	739
QY	864	GCACAAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTCACTTCAAGAAGAAAGCAGT	923
Db	740	RLFNLDVPESRRCFVKVRAYRSEFLPSEQIQGVVISVINLEPRTGFLSNPRAWGRFDSV	799
QY	924	GTGCAGGTGTACCATCTCCCAGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCAT	983
Db	800	ITGPNAGACVPAFCDDQSPDAYSAYVLASLAGEELQAVESSPKFNPNAIGVPQPYLNKLN	859
QY	984	GCTCCTTGGGTGGTGGCCAACTTGTCAATTATACCTCCAAGGACAACAGAGTGGTACATAA	1043
Db	860	RRTDHEDPRVKKTAFAQISMAKPRPNSAEESNGPIYAFENLRACEEAPPSAAHFRFYQIEG	919
QY	1044	GGCTAAAACAGAGTTGTCAACCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCA	1103
Db	920	DRYDYNTPVFNEDDPMSWTEGYLAWWPKPMEFRACYIKVKIVGPLEVNVRSRNMGGTHRR	979
QY	1104	GGGGTCTGGCACCTTCCAGGACCCTACTCTGCCTTTGCCCTTGTGGGATTTCCTTTAAAG	1163
Db	980	TVGKLYGIRDVRSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLKVVIPQGSRRASVN	1039

```

Qy      1164 CAACCGTGTCGGGCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAAT 1223
          : : : : : : : | : : : | : : : : | : : | : |
Db      1040 PMLHEYLVNHLPLAVNNDTSEYTMLAPLDPLGHNYGIYTVTDQDPRTAKEIALGRCFDGT 1099

Qy      1224 C 1224
          :
Db      1100 S 1100

```

RESULT 24

US-10-184-644-524

; Sequence 524, Application US/10184644

; Publication No. US20030044930A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

```

; APPLICANT:  Smith,Victoria

```

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C227

: CURRENT APPLICATION NUMBER: US/10/184,644

: CURRENT FILING DATE: 2002-06-28

```

; CURRENT FILING DATE: 2002 06 20
; Prior Application removed - See File Wrapper or Palm

```

: NUMBER OF SEO ID NOS: 612

: SEQ ID NO 524

```

; SEQ ID NO 324
:   LENGTH: 686

```

```
; LENGTH: 6
; TYPE: PRT
```

```

; TYPE: PRI
; ORGANISM: Homo Sapien

```

US-10-184-644-524

Query Match 2.4%; Score 38.2; DB 15; Length 686;
Best Local Similarity 11.9%; Pred. No. 0.15;
Matches 47; Conservative 108; Mismatches 241; Indels 0; Gaps 0;

Qy 944 AGTCAGAGACCCAGTAATCAGAGCAGCTAATGGGAGGCATGCTCCTTGGGTGGTGCCAA 1003
| | : : ||: : : : : : : :|| : : :| : : :
Db 178 ALFCOOLWRMGMLGTRVLSLVLFYKAYHFVVVAGAHWLVMTFWLVAQQSDIIDSTCHW 237

Qy 1004 CTTGTCATTATACCTCCAAGGACAACAGAGTGGTACATAAGGCTAAAACAGAGTTGTCAA 1063
 : :: | : : :: : : : : : : : : : : : : : :
Db 238 RLENLLVGAVYILCYLSFWDSPSRNRMVTFYMVLLENIIILLLATDFLQGASWTSLQTI 297

Qy 1064 CCTGTCCAGGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGG 1123
|::: : : : : : : || :: | ||::| :: | : ::
Db 298 AGVLSGFLIGSVSLVIYYSL LHPKSTDIWQGLRKSCGIAGGDKTERRDSPRATDLAGKR 357

Qy 1124 ACCCTACTCTGCCTTTGCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCTGGGGCCTTGGT 1183

Qy 1064 CCTGTCCAGGGGCAACTGGGATGGGGTAGGGCTGGGAGCAGGGGTCTGGCACCTTCCAGG 1123

```

      |::: : : : : : : : | | ::: | ||::| ::: |: ::
Db      298 AGVLSGFLIGSVSLVIYYSLLHPKSTDWQGLRKSCGIAGGDKTERRDSPRATDLAGKR 357

Qy      1124 ACCCTACTCTGCCTTTGCCCTTGTGGGATTTTCCTTTAAAGCAACCGTGTCTGGGCCTTGGT 1183
      :: | : : | : : | : : | : : : : : : : ||
Db      358 TESSGSCQGASYEPTILGKPPTPEQVPPEAGLGTQVAVEDSFLSHHHWLWVKLALKTGNV 417

Qy      1184 GGAACATCAAATCATGCCAGCAGAAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGT 1243
      :: | : : : | : : : | : : : : : : :
Db      418 SKINAAFGDNSPAYCPPAWGLSQQDYLRKALSAQQELPSSSRDPSTLENSSAFEGVPKA 477

Qy      1244 ACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGCCTGGGGGGGSC 1303
      : : : : : : | | | | : : | : | : | :
Db      478 EADPLETSSYVSFASDQQDEAPTQNPAATQGEQTPKEGADAVSGTQKGKTGGQQRGEGQ 537

Qy      1304 SGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAGG 1339
      :: : : | : | : | : | : : | : : |
Db      538 QSSTLYFSATAEVATSSQQEGSPATLQTAHSGRRLG 573

```

RESULT 26

US-10-027-632-147012

; Sequence 147012, Application US/10027632

; Publication No. US20020198371A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 147012

; LENGTH: 744

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-147012

```

Query Match          2.3%;   Score 36.6;   DB 13;   Length 744;
Best Local Similarity 58.9%;   Pred. No. 0.54;
Matches 63;   Conservative 0;   Mismatches 44;   Indels 0;   Gaps 0;

```

Qy 1396 GCCATGCATTTGGCATTGAATACAATCTGGTGACTTGTCTGGCTGCCAATAGAACCTAG 1455
 | | | | | | | | | | | | | | | | | | | | | |
 Db 464 GTCAAATATTTTACACTTAACATGCTGTACATAAAATATCTCCAAATAACTTCAA 523

Qy 1456 TACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAGTC 1502
 | | | | | | | | | | | | | | | | | | | | | |
 Db 524 TTCTAACCTGTAACCAAATGTGAAATCCCTGGGAAGACTGGAAAGTC 570

RESULT 27

US-10-027-632-147012
 ; Sequence 147012, Application US/10027632
 ; Publication No. US20030204075A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; TITLE OF INVENTION: Polymorphisms in the Human Genome
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 147012
 ; LENGTH: 744
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-027-632-147012

Query Match 2.3%; Score 36.6; DB 16; Length 744;
 Best Local Similarity 58.9%; Pred. No. 0.54;
 Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 1396 GCCATGCATTTGGCATTGAATACAATCTGGTGACTTGTCTGGCTGCCAATAGAACCTAG 1455
 | | | | | | | | | | | | | | | | | | | | | |
 Db 464 GTCAAATATTTTACACTTAACATGCTGTACATAAAATATCTCCAAATAACTTCAA 523

Qy 1456 TACCAAAGTGAAATCTTGAGGAAAATCCCTGGAAAGAGTGGAAAGTC 1502
 | | | | | | | | | | | | | | | | | | | | | |
 Db 524 TTCTAACCTGTAACCAAATGTGAAATCCCTGGGAAGACTGGAAAGTC 570

RESULT 28

US-10-377-079-11/c

```
; Sequence 11, Application US/10377079
; Publication No. US20030236395A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,
; TITLE OF INVENTION: Antibodies and Methods
; FILE REFERENCE: P-LJ 3611
; CURRENT APPLICATION NUMBER: US/10/377,079
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/389,956.
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1455)
US-10-377-079-11
```

```
Query Match          2.3%; Score 36.2; DB 16; Length 2236;
Best Local Similarity 50.3%; Pred. No. 1.4;
Matches 83; Conservative 2; Mismatches 80; Indels 0; Gaps 0;
```

```
Qy      1174 GGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGAT 1233
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      713 GTGCCTTGCTGGATCCTCTGCGCCGCGCAGATGCCGTAGGCCAGGCCGGGCACAGTACTG 654

Qy      1234 GTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGTAAGTGC 1293
          | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      653 GTGCAGAGGCACACCTCGCGAGGCAGGTCCCGCAGCCACTCCGGCAGCTCCGGCGGGGGC 594

Qy      1294 CTGGGGGGGSCSGGGGCTCCTGTACTTCTAAGGCAGGCTCTGGGAG 1338
          | | | | : : | | | | | | | | | | | | | | | |
Db      593 GCGGCGGCCGCGAGCGCTGCTGGTGCCCAAGCCGGCGCAGCGAG 549
```

RESULT 29

US-10-377-079-9/c

```
; Sequence 9, Application US/10377079
; Publication No. US20030236395A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,
; TITLE OF INVENTION: Antibodies and Methods
; FILE REFERENCE: P-LJ 3611
; CURRENT APPLICATION NUMBER: US/10/377,079
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/389,956.
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2488
; TYPE: DNA
```


Db 207 TGGGCCGGGGTCTTACCCATTTTTTTTGGGTTCGGTTGGAACAAGGGCCCTTTTCCCCT 148

QY 153 CCTCTGTGCCAGCCTTTCTCCCAGCATTCCTYTCTGGC 190
| | | | | | | | | | | | | | | | | | | | | |

Db 147 TTTGTTTCCCTGCCTCTGTAATAGCCTTTTTTCCCGGCC 110

US-09-867-701-2637

; Patent No. US20020132237A1

; APPLICANT: Aglate, Paul A.

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

; CURRENT APPLICATION NUMBER

```
; NUMBER OF SEQ ID NOS: 10912
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
```

; SEQ ID NO 2637

; LENGTH: 257

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-867-701-2637

Best Local Similarity 58.5%; Pred. No. 0.64;

Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 780 CTGATTTCTGCTCTCCCTTCTCTTGACTCGCCCACCACTGTCTGTGTAGATGGAGAAG 839
 ||||| ||| ||||| | | ||| | | ||| ||| ||| ||| |||
Db 12 CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCTCCCTCTGGGAATTGGGCAG 71

λ_1

Db 72 CCCTGGGCAGTTGTACTCATGGGGGCTTAAGATGCAGCTACCTCAG 117

RESULT 32

US-09-880-107-893/c

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; APPLICANT: Gene Logic, Inc.

10; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880,107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

```
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 893
;   LENGTH: 330
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA411813
US-09-880-107-893
```

Qy	780	CTGATTTCTGCTCTCCCTTCCTTGACTCGCCACCACTGTCTGTGTAGATGGAGAAG	839
Db	282	CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCTCACCTCTGGGAATTGGGCAG	223
Qy	840	GCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG	885
Db	222	CCCTGGGCAGTTGTACTCATGGGGCTTAAGATGCAGCTACCTCAG	177

US-09-867-701-10557/c

; Patent No. US20020132237A1

; APPLICANT: Aglate, Paul A.

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

; CURRENT APPLICATION NUMBER: US/09/867,701

; NUMBER OF SEQ ID NOS: 10912

; SEO ID NO 10557

; TYPE: DNA

; FEATURE:

; LOCATION: (1) ... (440)

US-09-867-701-10557

Qy 780 CTGATTTCTGCTCTCCCTTCTTGACTCGCCACCACTGTCTGTGTAGATGGAGAAG 839
 |||| | | ||||| | | | ||| | | | ||| | | | ||| | | |
 Db 261 CTGAAGCCGGGGCTCCCCCTGCCTGCCTCTCTCTCCTCCTCCCTCTGGGAATTGGGCAG 202

Qy 840 GCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
 | | | | | | | | | | | | | | | | | | | |
 Db 201 CCCTGGGCAGTTGTACTCATGGGGCTTAAGATGCAGCTACCTCAG 156

RESULT 34

US-10-184-644-348

; Sequence 348, Application US/10184644

; Publication No. US20030044930A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C227

; CURRENT APPLICATION NUMBER: US/10/184,644

; CURRENT FILING DATE: 2002-06-28

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 348

; LENGTH: 777

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-184-644-348

Query Match 2.3%; Score 35.6; DB 15; Length 777;

Best Local Similarity 9.4%; Pred. No. 1.2;

Matches 41; Conservative 138; Mismatches 256; Indels 0; Gaps 0;

Qy 308 GGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTTAAAGCTG 367
 :: :: |: :: : ||:: : : : | : :
 Db 103 KKIYWPAAKERVELCKLAGKDANTECANFIRVLQPYNKTHIYVCGTGAFHPICGYIDLGV 162
 Qy 368 CCCTGGAGCCGGACAGGCCACTAGAAAATTCAC TTGCATTTGCTTCCTGCTAGCCATGGG 427
 : : : : | | : : : : : : : | |
 Db 163 YKEDIIFKLDTHNLESGR LKCPDPQPFASVMTDEYLYSGTASDFLGKDTAFTRSLGPT 222
 Qy 428 TGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCT 487
 :: : : : : | | : : : : : : : :
 Db 223 HDHHYIRTDISEHYWLNGAKFIGTFFIPDTYNPDDDKIYFFRESSQEGSTSDKTILSRV 282
 Qy 488 GAGCTCCCTGGAGCAAGGTTCTGGTACGCGGCACAGAGGCTCGGCACAGCTTAGGTGTCCT 547
 |::|: : | :: : || : : : : : : : |:: : :
 Db 283 GRVCKNDVGGQ RSLINKWTTFLKARLICSIPGSDGADTYFDELQDIYLLPTRDERNPVVY 342
 Qy 548 GCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGAT 607
 | : | | : : : |::: : : | : : : : : | | : |
 Db 343 GVFTTTSSIFKGSACVYSMADIRAVFNPGPYAHKESADHRWVQYDGRIPYPRPGTCPSKT 402

Qy 608 TGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAG 667
 :: : | : :: : :::: : :: : | :: :: | :: :: : |
 Db 403 YDPLIKSTRDFPDDVISFIKRHSVMYKSVYPVAGGPTFKRINVDYRLTQIVVDHVIAEDG 462
 Qy 668 GGTGAAAGTGGACAGTCTGTAAACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGG 727
 :: | | : : : : : : : : : : : : : : : | :
 Db 463 QYDVMFLGTDIGTVLKVVVISKEKWNMEEVVLEELQIFKHSSIILNMELSLKQQQLYIGS 522
 Qy 728 CAGAGCCTGGACATT 742
 : | : : | : | :
 Db 523 RDGLVQLSLHRCPTY 537

RESULT 35

US-10-184-634-348

; Sequence 348, Application US/10184634

; Publication No. US20030068684A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C217

; CURRENT APPLICATION NUMBER: US/10/184,634

; CURRENT FILING DATE: 2002-06-28

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 348

; LENGTH: 777

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-184-634-348

Query Match 2.3%; Score 35.6; DB 15; Length 777;

Best Local Similarity 9.4%; Pred. No. 1.2;

Matches 41; Conservative 138; Mismatches 256; Indels 0; Gaps 0;

Qy 308 GGATTCCAGGGCTGGGTAGGATCGGACAGGGCACTCCCATTTGGCTCCTCAGTTAAAGCTG 367
 :: :: | : :: : || :: : : : : | : : : :
 Db 103 KKIYWPAAKERVELCKLAGKDANTECANFIRVLQPYNKTHIYVCGTGAFHPICGYIDLGV 162
 Qy 368 CCCTGGAGCCGGACAGGCCACTAGAAAATTCACTTGCATTTGCTTCCTGCTAGCCATGGG 427
 : : : : | | : : : | : : : : | | |
 Db 163 YKEDIIFKLDTHNLESGRLKCPFDPPQPFASVMTDEYLYSGTASDFLGKDTAFTRSLGPT 222
 Qy 428 TGAGCTGCCCTTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCT 487
 :: : : : : | | : : : : : : : :

Db 223 HDHHYIRTDISEHYWLNKAKFIGTFFIPDTYNPDDDKIYFFRESSQEGSTSDKTILSRV 282

Qy 488 GAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCT 547
|::|: : | :: : || : : ::||: : : |:: :::

Db 283 GRVCKNDVGGQRLINKWTTFLKARLICSIPGSDGADTYFDELQDIYLLPTRDERNPVVY 342

Qy 548 GCATGTGTCTACAGCGTCAGGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGAT 607
|: | |: : :|::: : :|: | : : : :| | : |

Db 343 GVFTTTSSIFKGSAVCVYSMADIRAVFNGPYAHKESADHRWVQYDGRIPIYPRPGTCPSKT 402

Qy 608 TGCCTTTTCTGAATGGGTGGGTGGGCCTGTGGGCTTTGGGTGTCTGTCCAGCAGATCAG 667
:: : | : :: : : :::: : : : : | :: :: : | :: :: : |

Db 403 YDPLIKSTRDFPDDVISFIKRHSVMYKSVYPVAGGPTFKRINVDYRLTQIVVDHVIAEDG 462

Qy 668 GGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTTCCTCCTCCTCCTCCTGCGCAGGG 727
::: ||: : :: : : : : : : : : : : : : : |:

Db 463 QYDVMFLGTDIGTVLKVVSISKEKWNMEEVVLEELQIFKHSSIILNMELSLKQQQLYIGS 522

Qy 728 CAGAGCCTGGACATT 742
:| : :|:::

Db 523 RDGLVQLSLHRCPTY 537

RESULT 36

US-10-087-192-1894/c

; Sequence 1894, Application US/10087192

; Publication No. US20020182586A1

; GENERAL INFORMATION:

; APPLICANT: Morris, David W.

; APPLICANT: Engelhard, Eric K.

; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: CANCER

; FILE REFERENCE: 529452000122

; CURRENT APPLICATION NUMBER: US/10/087,192

; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 09/747,377

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: US 09/798,586

; PRIOR FILING DATE: 2001-03-02

; NUMBER OF SEQ ID NOS: 2059

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 1894

; LENGTH: 35143

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-087-192-1894

Query Match 2.3%; Score 35.6; DB 13; Length 35143;

Best Local Similarity 46.7%; Pred. No. 11;

Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 646 GGTGTGTCTGTCCAGCAGATCAGGGTCAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTT 705
|| | |||| | | || | | |||| | | || | |

Db 25274 GGCAGGCTGTTCTCTGGTTCCAACACTTGTCCACAGGATCTCTAAAGACCCAGGAATGG 25215

Qy 706 CTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCG 765
|| | | | || || || || | | |||| | | || |

Db 25214 GGGCTATTGCCAGGGGTTAGAAGAGAACCAGGTCCCAAGGGCATGGTGGGCGGGCAGATG 25155
 Qy 766 CTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCTTG 825
 || | | || | |||| | || | |||| || || | ||||
 Db 25154 GTTCCAGAGCCTTAGAGATTCATAGGTTCTTCCTCCTCCACCAGCTGCTCCGAGGGCCTG 25095
 Qy 826 TG TAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885
 || || || || || | || | || | || | || || || ||
 Db 25094 TGGGGAGGGACAAGGGTGGGATGCTGGAGCACCAGGGCTGCAGCAAGGGCCTTAGCTAAG 25035
 Qy 886 CT 887
 ||
 Db 25034 CT 25033

RESULT 37

US-09-563-728A-36/c

; Sequence 36, Application US/09563728A
 ; Publication No. US20030078216A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MacLeod, Alan R
 ; APPLICANT: Li, Zoumei
 ; APPLICANT: Besterman, Jeffrey M
 ; TITLE OF INVENTION: Inhibition of Histone Deacetylase
 ; FILE REFERENCE: 106101.229
 ; CURRENT APPLICATION NUMBER: US/09/563,728A
 ; CURRENT FILING DATE: 2000-05-03
 ; PRIOR APPLICATION NUMBER: 60/132,287
 ; PRIOR FILING DATE: 1999-05-03
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 36
 ; LENGTH: 122186
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-563-728A-36

Query Match 2.3%; Score 35.6; DB 10; Length 122186;
 Best Local Similarity 46.7%; Pred. No. 22;
 Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 646 GGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGTAACAACAGTGAGTCGTTTC 705
 || | |||| | || | |||| || || || ||
 Db 107424 GGCAGGCTGTTCTCTGGTTCCAACACTTGTCCACAGGATCTCTAAAGACCCAGGAATGG 107365
 Qy 706 CTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAACATGCCCTGCCTGAAGCCG 765
 || | | | || || || || || || || || || ||
 Db 107364 GGGCTATTGCCAGGGGTTAGAAGAGAACCAGGTCCCAAGGGCATGGTGGGCGGGCAGATG 107305
 Qy 766 CTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACTCGCCCACCACCTGTCTTG 825
 || | | || | |||| | || | |||| || || | ||||
 Db 107304 GTTCCAGAGCCTTAGAGATTCATAGGTTCTTCCTCCTCCACCAGCTGCTCCGAGGGCCTG 107245
 Qy 826 TG TAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCACAAAATGGAATGAACACTG 885

Db 107244 TGGGGAGGGACAAGGGTGGGATGCTGGAGCACCAGGGCTGCAGCAAGGGCCTTAGCTAAG
107185

Qy 886 CT 887

Db 107184 CT 107183

RESULT 38

US-10-142-426-358

; Sequence 358, Application US/10142426

; Publication No. US20040048333A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C224

; CURRENT APPLICATION NUMBER: US/10/142,426

; CURRENT FILING DATE: 2002-05-09

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 358

; LENGTH: 1049

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-142-426-358

Query Match 2.3%; Score 35.4; DB 13; Length 1049;

Best Local Similarity 4.3%; Pred. No. 1.7;

Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

Qy 208 ACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267

Db 1 MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy 268 ATGGCATCGTTGGCCAAATTGGTGAACGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327

Db 61 GGIPTNTTNLTTLTINHIPDISPASFHRLDHLVEIDFRNCNCVPIPLGSKNNMCIKRLQIKP 120

Qy 328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387


```

      : :| : : : : : : : : : : : : : : : :
Db      121 RSFSGLTYLKSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCACCTTGCACTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
      : : : : : : | : : : : | : | : : : : :
Db      181 QNCYRNP CYVSY SIEKDAFLNLT KLKVL SLKDN NVTA VPTVLPSTLT ELYLYNNMIAKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
      : : : : : : : | : : : : : : : : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy      508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCA 567
      : : : : : : : | : : : : : : : : : :
Db      301 SLQHVPPRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRASM 360

Qy      568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627
      | : : : : : : : : : : : : : : : | : | : :
Db      361 NLSQAFSSLSKSLKILRIRGYVFKELKSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420

Qy      628 GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
      : : : : : : : : | | : : : : : : : : | : :
Db      421 KRLKVIDLSVNKISPSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480

Qy      688 AACACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC 747
      | : : : : : : : | : : : : : : : : : : :
Db      481 EASFMSVNESCYKYGTLDLSKNSIFFVKSSDFQHLSFLKCLNLSGNLISQTLNGSEFQP 540

Qy      748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACT 807
      : : : : : : : | : : : : : : : : : | : : :
Db      541 LAELRYLDFSNNRDLHLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKV LQK 600

Qy      808 CGCCACCACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
      : : : : : : : : : : : : : : : | : : : : :
Db      601 LMMNDNDISSSTSRTMESESLRTEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKN 660

Qy      868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
      : : : : : : : | : : : : : : : |
Db      661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQC 697

```

RESULT 39

US-10-123-155-358

; Sequence 358, Application US/10123155

; Publication No. US20030068794A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

```

; APPLICANT:  Stewart, Timothy A.
; APPLICANT:  Tumas, Daniel
; APPLICANT:  Watanabe, Colin K
; APPLICANT:  Wood, William
; APPLICANT:  Zhang, Zemin
; TITLE OF INVENTION:  SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION:  ACIDS ENCODING THE SAME
; FILE REFERENCE:  P3330R1C30
; CURRENT APPLICATION NUMBER:  US/10/123,155
; CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 358
;   LENGTH: 1049
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-123-155-358

```

```

Qy      748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTCTGCTCTCCCCTTCCTTGACT 807
          : : : : : : | : : : : : : | : : : : : :
Db      541 LAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKVQLQK 600

Qy      808 CGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
          : : : : : : : : : : : : | : : : : : :
Db      601 LMMNDNDISSSTSRTMESESLRTLEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKN 660

Qy      868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
          : : : : : : | : : : : : :
Db      661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFWSKKLQC 697

```

RESULT 40

US-10-146-731-358

; Sequence 358, Application US/10146731

; Publication No. US20030129692A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C323

; CURRENT APPLICATION NUMBER: US/10/146,731

; CURRENT FILING DATE: 2002-05-15

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 358

; LENGTH: 1049

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-146-731-358

Query Match 2.3%; Score 35.4; DB 15; Length 1049;

Best Local Similarity 4.3%; Pred. No. 1.7;

Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```

Qy      208 ACACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
          : : : | : : : : : : : : : : : : : : : :
Db      1 MVFPMWTLKRQILILFNIIILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60

```

Qy	268	ATGGCATCGTTGGCCAAATTGGTGAACGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG	327
Db	61	GGIPTNTTNLTLTINHIPDISPASFHRLDHLVEIDFRCNCVPIPLGSKNNMCIKRLQIKP	120
Qy	328	ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA	387
Db	121	RSFSGLTLYLKSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG	180
Qy	388	CTAGAAAATTCACTTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC	447
Db	181	QNCYYRNPCYVSYSIEKDAFLNLTCLKVLSLKDNVTVAVPTVLPSTLTELYLYNNMIAKI	240
Qy	448	CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT	507
Db	241	QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN	300
Qy	508	CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTACAGCGTCA	567
Db	301	SLQHVPPrWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFEQVYRASM	360
Qy	568	GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG	627
Db	361	NLSQAFSSLSKSLKILRIRGYVFKEKLSFNLSPLHLNLQNLEVLDLGTNFIKIANLSMFKQF	420
Qy	628	GTGGGCCTGTGGGCTTTGGGTGTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT	687
Db	421	KRLKVIDLSVNKISPSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK	480
Qy	688	AACAACAGTGAGTCGTTTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC	747
Db	481	EASFMSVNESCYKYQOTLDLSKNSIFFVKSSDFQHLSFLKCLNLSGNLISQTLNGSEFQP	540
Qy	748	ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACT	807
Db	541	LAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKVQLK	600
Qy	808	CGCCCAACCACTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC	867
Db	601	LMMNDNDISSSTSRTMESESLRTELEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKN	660
Qy	868	AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC	904
Db	661	SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQC	697

```

; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 358
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-472-358

```

```

Query Match          2.3%; Score 35.4; DB 15; Length 1049;
Best Local Similarity 4.3%; Pred. No. 1.7;
Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```

```

Qy      208 ACACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
      :: : | ::          ::      : :          :: : :: :: ::
Db      1 MVFPMWTLKRQILILFNILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
      | | : : : : : : : : : : : : | : : : | :: :
Db      61 GGIPTNTTNLTLTINHIPDISPASFHRLDHLVEIDFRNCNCPVPIPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
      : :| :: : : : : : : : : : : : : : : :
Db      121 RSFSGLTLYKSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCACTTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
      : :: : : : | : : : : : | : | : : : : :
Db      181 QNCYRNP CYVSYSIEKDAFLNLT KLKVL SLKDN NVTA VPTVLPSTLTELYLYNNMIAKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
      :: : : : : | : : : : : : : : : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy      508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCA 567
      : :: : : : : : : | : : : : : : : : : :
Db      301 SLQHVPFRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRASM 360

Qy      568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627
      | :: : : : :| :: : : : : : : : : : | : | : :
Db      361 NLSQAFSSLKSLKILRIRGYVFKEKLSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420

Qy      628 GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
      : :: : : : : :: | | : : : : : : : : | : :

```

```

Db      421 KRLKVIDLSVNKISPSGDSSEVGFCNSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480
Qy      688 AACAAACAGTGAGTCGTTCCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAC 747
      |: ::: : ::| | : : :::: :| : :
Db      481 EASFMSVNESCYKYGQTLDSLKNSIFFVKSSDFQHLNFLKCLNLSGNLISQTLNGSEFQP 540
Qy      748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTCTGCTCTCCCCTTCCTTGACT 807
      : : : : : :| : : : : : : |:: | : : :
Db      541 LAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKVQLK 600
Qy      808 CGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
      : : : : : : : : : : : : : :
Db      601 IMMNDNDISSSTSRTMESESLRLEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKN 660
Qy      868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
      : : : : : : |: : : : :
Db      661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQC 697

```

RESULT 42

US-10-141-761-358

; Sequence 358, Application US/10141761

; Publication No. US20030148432A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C198

; CURRENT APPLICATION NUMBER: US/10/141,761

; CURRENT FILING DATE: 2002-05-08

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 358

; LENGTH: 1049

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-141-761-358

Query Match 2.3%; Score 35.4; DB 15; Length 1049;

Best Local Similarity 4.3%; Pred. No. 1.7;

Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

Qy 208 ACACCGTGTGTTCTGCCTATTGTCTGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
 :: : | :: :: :: : :: : :: : ::
 Db 1 MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy 268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
 | | : : : : : : : : | : : | :: :
 Db 61 GGIPTNTTNLTTLTINHIPDISPASFHRLDHLVEIDFRCNCVPIPLGSKNNMCIKRLQIKP 120

Qy 328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
 : :| :: : : : : : : : : :
 Db 121 RSFSGLTLYKSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy 388 CTAGAAAATTCACTTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
 : :: : : : | : : : : | : | : : : :
 Db 181 QNCYYRNP CYVSY SIEKDAFLNLT KLKVL SLKDN NVTA VPTVLP STLTELYLYNNMI AKI 240

Qy 448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
 :: : : : | : : : : : : : :
 Db 241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy 508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCA 567
 : :: : : : : : | : : : : : : : : :: ::
 Db 301 SLQHVPPRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRASM 360

Qy 568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627
 | :: : : : : | :: : : : : | : | : : :
 Db 361 NLSQAFSSLKSLKILRIRGYVFKEKLSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420

Qy 628 GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
 : :: : : : : : | | : : : : : : : : | : : :
 Db 421 KRLKVIDLSVNKISPSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480

Qy 688 AACACAGTGAGTCGTTCCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC 747
 | : :: : : : : | | : : : : : : : | : : :
 Db 481 EASFMSVNESCYKYQTLDLSKNSIFFVKSSDFQHL SFLKCLNLSGNLISQTLNGSEFQP 540

Qy 748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACT 807
 : : : : : : | : : : : : : : : | : : | : : :
 Db 541 LAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKV LQK 600

Qy 808 CGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
 : : : : | : : : : : : : : : : : : : :
 Db 601 LMMNDNDISSSTSRTMESES LRTLEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKN 660

Qy 868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
 : : : : : : | : : : : : : |
 Db 661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQC 697

RESULT 43

US-10-142-885-358

; Sequence 358, Application US/10142885

; Publication No. US20030157604A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

```

; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 358
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-885-358

```

```

Query Match          2.3%; Score 35.4; DB 15; Length 1049;
Best Local Similarity 4.3%; Pred. No. 1.7;
Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```

```

Qy      208 ACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
      :: : | ::          :: : : :          :: : : : : :
Db      1 MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
      | | : : : : : : : : | : : | : :
Db      61 GGIPTNTNLTLTLINHIPDISPASFHRDLHLVEIDFRNCNVPIPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
      : : | : : : : : : : : : : :
Db      121 RSFSGLTLYKSLYLDGNQLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCAC TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
      : : : : : | : : : : | : : | : : :
Db      181 QNCYYRNP CYVSY SIEKDAFLNLTKLVLSLKDNNVTAVPTVLPSTLT ELYLYNNMI AKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
      :: : : : | : : : : : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy      508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCA 567
      : : : : : : : | : : : : : : :
Db      301 SLQHVPPRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRASM 360

Qy      568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627

```



```

      | :: : : : :|:: : : : : | : | : :
Db      361 NLSQAFSSLKSLKILRIRGYVFKELKSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420

Qy      628 GTGGGCCTGTGGGCTTTGGGTGTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
      : :: : : : :| | : : : :| : :
Db      421 KRLKVIDLSVNKISPSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480

Qy      688 AACAACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC 747
      |: :: : : :| | : : : : :| : :
Db      481 EASFMSVNESCYKYGQTLDSLKNSIFFVKSSDFQHLSFLKCLNLSGNLISQTLNGSEFQP 540

Qy      748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACT 807
      : : : : : :| : : : : :| : :
Db      541 LAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKVQLK 600

Qy      808 CGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
      : : : : : : : : : :| : : : :
Db      601 LMMNDNDISSSTSRTMESESLRTEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKN 660

Qy      868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
      : : : : : :| : : : :
Db      661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFWSWKKLQC 697

```

RESULT 44

US-10-158-790-358

; Sequence 358, Application US/10158790

; Publication No. US20030180879A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C448

; CURRENT APPLICATION NUMBER: US/10/158,790

; CURRENT FILING DATE: 2002-05-30

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 358

; LENGTH: 1049

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-158-790-358

Query Match 2.3%; Score 35.4; DB 15; Length 1049;
Best Local Similarity 4.3%; Pred. No. 1.7;
Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```
Qy      208 ACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
          :: : | ::          ::      :: :          :: : :: : :: :
Db      1  MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
          |  |  : : : : : : : : : : | : : : | :: :
Db      61 GGIPTNTTNLTLTINHIPDISPASFHRLDHLVEIDFRCNCVPIPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
          : : | :: : : : : : : : : : : : : : :
Db      121 RSFSGLTLYKSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCACCTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
          : : : : : | : : : : | : | : : : :
Db      181 QNCYYRNPCYVSYIEKDAFLNLTCLKVLSLKDNNVTAVPTVLPSTLTELYLYNNMIAKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
          :: : : : | : : : : : : : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy      508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTTACAGCGTCA 567
          : :: : : : : : : | : : : : : : : : :
Db      301 SLQHVPPRWFKNINKLQELDLSQNF LAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRASM 360

Qy      568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627
          | :: : : : : | :: : : : : | : | : : :
Db      361 NLSQAFSSLKSLKILRIRGYVFELKSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420

Qy      628 GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
          : :: : : : : : : | | : : : : : : : | : :
Db      421 KRLKVIDLSVNKISPSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480

Qy      688 AACACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC 747
          | : :: : : : : | : : : : : : : : :
Db      481 EASFMSVNESCYKYGTLDLSKNSIFFVKSSDFQHLSFLKCLNLSGNLISQTLNGSEFQP 540

Qy      748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCTTGA 807
          : : : : : : | : : : : : : : : : | : : :
Db      541 LAELRYLDFSNRLDLLHSTA FEELHKLEVLDISSNSHYFQSEGI THMLNFTKNLKV LQK 600

Qy      808 CGCCCAACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
          : : : : : : : : : : : : : : : : :
Db      601 LMMNDNDISSSTSRTMESESLRTEFRGNHLDVLWREGDNRYLQLFKNLLKLEELD ISKN 660

Qy      868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
          : : : : : : : | : : : : |
Db      661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQC 697
```

RESULT 45

US-10-137-871-358

```

; Sequence 358, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 358
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-871-358

```

```

Query Match          2.3%; Score 35.4; DB 16; Length 1049;
Best Local Similarity 4.3%; Pred. No. 1.7;
Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```

```

Qy      208 ACACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
      :: : | ::          ::  :: :          :: : :: :: :: ::
Db      1  MVFPMWTLKRQILILFNIIILSKLLGARWFPKTLPDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
      |  |  : : : : : : : : : : | :  :  | :: :
Db      61 GGIPTNTTNTLTLTINHIPDISPASFHRDLHLVEIDFRCNCVPIPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
      : :|  :: :  :          :          : :  :
Db      121 RSFSGLTLYLKSLYLDGNQLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCACTTGCAATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
      :  :: :  :  | :  : : : | : :|  : :  :
Db      181 QNCYYRNPCYVSYSEIKDAFLNLTCLKVLSLKDNNVTAVPTVLPSTLTLEYLYNNMIAKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
      ::          : :  :|          :  :  :          : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

```


; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 358
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-923-358

Query Match 2.3%; Score 35.4; DB 16; Length 1049;
Best Local Similarity 4.3%; Pred. No. 1.7;
Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```
Qy      208 ACACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
          :: : | ::          :: : :          :: : : : : : :
Db      1  MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPHNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
          | | : : : : : : : : : : : : | : : : : :
Db      61 GGIPTNTTNLTTLTINHIPDISPASFHRLDHLVEIDFRNCNCPVPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
          : : | : : : : : : : : : : : : : : : :
Db      121 RSFSGLYLKSLEYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCACCTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
          : : : : : : | : : : : : : | : : : : :
Db      181 QNCYYRNP CYVSYSIEKDAFLNLT KLKVL SLKDN NVTA VPTVLPSTLT ELYLYNNMI AKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
          :: : : : : : | : : : : : : : : : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy      508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCA 567
          : : : : : : : : : | : : : : : : : : : :
Db      301 SLQHVP PRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFN FELQVY RASM 360

Qy      568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627
          | :: : : : : : | :: : : : : : | : : : : :
Db      361 NLSQAFSSLSKSLKILRIRGYVFKEKLSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420

Qy      628 GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
          : : : : : : : : | | : : : : : : : : | : :
Db      421 KRLKVIDLSVNKISPSGDSSEVGFC SNARTSVESYEPQVLEQLHYFRYDKYARSCR FKNK 480

Qy      688 AACACAGTGAGTCGTTCCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC 747
          | : :: : : : : | : : : : : : : : : : :
Db      481 EASFMSVNESCYKYGTLDLSKNSIFFVKSSDFQHLSFLKCLNLSGNLISQTLNGSEFQP 540

Qy      748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTCTGCTCTCCCCTTCCTTGACT 807
          : : : : : : : | : : : : : : : : : : :
Db      541 LAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKV LQK 600

Qy      808 CGCCACACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
          : : : : : : : : : : : : : : : : : :
Db      601 IMMNDNDISSSTSRTMESES LRTLEFRGNHLDVLWREGDNRYLQLFKNLLKLEELD ISKN 660

Qy      868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
          : : : : : : : | : : : : : |
```

RESULT 47

US-10-141-756-358

; Sequence 358, Application US/10141756

; Publication No. US20030207359A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C200

; CURRENT APPLICATION NUMBER: US/10/141,756

; CURRENT FILING DATE: 2002-05-08

; Prior Apploication removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 358

; LENGTH: 1049

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-141-756-358

Query Match 2.3%; Score 35.4; DB 16; Length 1049;

Best Local Similarity 4.3%; Pred. No. 1.7;

Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```

Qy      208 ACACCGTGTGTTCTGCCTATTGTCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
          :: : | ::          ::      :: :          :: : :: :: ::
Db      1  MVFPMWTLKRQILILFNILISKLLGARWFPKTLPCDVTLDPKHNHIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
          | | : : : : : : : : : : | : : : | :: :
Db      61 GGIPTNTNLTLTINHIPDISPASFHRDLHLVEIDFRNCVPIPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
          : : | : : : : : : : : : : : : : : :
Db      121 RSFSGLTLYLSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYL 180

Qy      388 CTAGAAAATTCACTTGCACTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
          : : : : : | : : : : | : : | : : :
Db      181 QNCYYRNP CYVSY SIEKDAFLNLTKLVLSLKDNNVTAVPTVLPSTLT ELYLYNNMIAKI 240

```

Qy	448	CAGAGGGAGCCAGAGAGGGCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGGCAAGGTT	507
Db	241	QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN	300
Qy	508	CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCA	567
Db	301	SLQHVPPrWfKNINKLQELDLsqNFLAKEIGDAKFLHFLPSLIQLDLsFNfELQVYRASM	360
Qy	568	GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG	627
Db	361	NLSQAFSSlKSLKILRIRGYVFkELKsFNLSPLHLNLQNLLEVLDLGTNfIKIANLSMfKQF	420
Qy	628	GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT	687
Db	421	KRLKVIDLSVNKISPSGDSSEVGfCSNARTSVESYEPQVLEQLHYFRYDKYARSCRfKNK	480
Qy	688	AACAACAGTGAGTCGTTCTCTCTCTCTCTGCGCAGGGCAGAGCCTGGACATTAAAAC	747
Db	481	EASFMSVNESCYKYQOTLDLSKNSIFFVKSSDFQHLsFLKCLNLsgNLISQTLNGSEfQP	540
Qy	748	ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCTTCCTTGACT	807
Db	541	LAELRYLDfSNRRLDLHSTAfEELHKLEVLDISSNSHYfQSEGITHMLNfTKNLKVLQK	600
Qy	808	CGCCCAccACCTGTCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC	867
Db	601	LMMNDNDISSSTsRTMESESRLTLEfRGNHLDVLWREGDNRYLQLfKNLLKLEELDIsKN	660
Qy	868	AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC	904
Db	661	SLSFLPSGVfDGMPPNLKNLSLAKNGLKSfSWKKLQC	697

US-10-141-759-358

; Sequence 358, Application US/10141759

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

: APPLICANT: Gurney, Austin L.

: APPLICANT: Sherwood, Steven

: APPLICANT: Smith-Victoria

: APPLICANT: Stewart, Timothy A.

: APPLICANT: Tumas Daniel

: APPLICANT: Watanabe Colin K

• APPLICANT: Wood William

: APPLICANT: Zhang Zemin

```
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C197
; CURRENT APPLICATION NUMBER: US/10/141,759
; CURRENT FILING DATE: 2002-05-08
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 358
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-759-358
```

```
Query Match          2.3%; Score 35.4; DB 16; Length 1049;
Best Local Similarity 4.3%; Pred. No. 1.7;
Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;
```

```
Qy      208 ACACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
      :: : | ::          :: :          :: : :: : ::
Db      1 MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
      | | : : : : : : : : | : : | :: :
Db      61 GGIPTNTNLTLTINHIPDISPASFHRLDHLVEIDFRCNCVPIPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
      : :| :: : : : : : : : : :
Db      121 RSFSGLTLYLKSLEYLDGNQLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCACCTGCAATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC 447
      : :: : : : | : : : : | : | : : :
Db      181 QNCYYRNP CYVSYSIEKDAFLNLTKLVLSLKDNNVTAVPTVLPSTLTELYLYNNMIAKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
      ::          : : : | : : : : : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy      508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCA 567
      : :: : : : : : | : : : : : : : : ::
Db      301 SLQHVPPRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRAS 360

Qy      568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627
      | :: : : : : | :: : : : : | : | : :
Db      361 NLSQAFSSLKSLKILRIRGYVFKEKLSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420

Qy      628 GTGGGCCTGTGGGCTTTGGGTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
      : :: : : : : : | | : : : : : : | : :
Db      421 KRLKVIDLSVNKISPSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480

Qy      688 AACAACAGTGAGTCGTTCCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC 747
      | : :: : : : : | : : : : : : : | : :
Db      481 EASFMSVNESCYKYGTLDLSKNSIFFVKSSDFQHLSFLKCLNLSGNLISQTLNGSEFQP 540

Qy      748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCTTCCTTGACT 807
      : : : : : : | : : : : : : : | : | : :
Db      541 LAELRYLDFSNRLDLLHSTA FEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKV LQK 600

Qy      808 CGCCACACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
```



```

      :      :|: : : : : : : : : : |: : : : :
Db      601 LMMNDNDISSSTSRTMESESRLTLEFRGNHLDVLRREGDNRYLQLFKNLLKLEELDISKN 660

Qy      868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
      : : : : : : : |: : : : : |
Db      661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQC 697

```

RESULT 49

US-10-140-805-358

; Sequence 358, Application US/10140805

; Publication No. US20030207417A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C176

; CURRENT APPLICATION NUMBER: US/10/140,805

; CURRENT FILING DATE: 2002-05-07

; Prior Apploication removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 358

; LENGTH: 1049

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-140-805-358

Query Match 2.3%; Score 35.4; DB 16; Length 1049;

Best Local Similarity 4.3%; Pred. No. 1.7;

Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;

```

Qy      208 ACACCGTGTGTTCTGCCTATTGTCTGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
      : : | : : : : : : : : : : : : : : : :
Db      1 MVFPMWTLKRQILILFNILISKLLGARWFPKTLPCDVTLDVPHNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAACGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
      | | : : : : : : : : : : : : : : : :
Db      61 GGIPTNTNLTTLTINHIPDISPASFHRLDHLVEIDFRCNCVPIPLGSKNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
      : :| : : : : : : : : : : : : : : : :

```



```
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C184
; CURRENT APPLICATION NUMBER: US/10/140,864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 358
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-864-358
```

```
Query Match          2.3%; Score 35.4; DB 16; Length 1049;
Best Local Similarity 4.3%; Pred. No. 1.7;
Matches 30; Conservative 183; Mismatches 484; Indels 0; Gaps 0;
```

```
Qy      208 ACACCGTGTGTTCTGCCTATTGTGCGAGATAAGGACACTCTGGCTAAAGGTACATCAGATA 267
      :: : | ::          ::      :          :: : :: :: ::
Db      1 MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60

Qy      268 ATGGCATCGTTGGCCAAATTGGTGAAGTGTATCTCACGAGGATTCCAGGGCTGGGTAGG 327
      | | : : : : : : : : : | : : : | :: :
Db      61 GGIPTNTTNLTLTINHIPDISPASFHRLDHLVEIDFRCNCVPIPLGSKNNMCIKRLQIKP 120

Qy      328 ATCGGACAGGGCACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCA 387
      : : | :: : : : : : : : : : : : :
Db      121 RSFSGLTYLKSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180

Qy      388 CTAGAAAATTCACATTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTTCTGAGTC 447
      : : : : : | : : : : | : | : : : :
Db      181 QNCYYRNPCYVSYSEKDAFLNLTCLKVLSLKDNNVTAVPTVLPSTLTELYLYNNMIAKI 240

Qy      448 CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTT 507
      :: : : : | : : : : : : : :
Db      241 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300

Qy      508 CGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGCTGATGTGTCTACAGCGTCA 567
      : :: : : : : : : | : : : : : : : :
Db      301 SLQHVPPRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRASM 360

Qy      568 GGTAAGGGGACCTCCACAGCAAAAAGCTAGGCTCTCTGATTGCCTTTTCTGAATGGGTGG 627
      | :: : : : : | : : : : : : : | : : :
Db      361 NLSQAFSSLKSLKILRIRGYVFKELKSFNLSPLHNLQNLEVLDTNFIKIANLSMFKQF 420

Qy      628 GTGGGCCTGTGGGCTTTGGGTTGTCTGTCCAGCAGATCAGGGTGAAAGTGGACAGTCTGT 687
      : :: : : : : : : | | : : : : : : : :
Db      421 KRLKVIDLSVNKISPSGDSSEVGFCNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480

Qy      688 AACACAGTGAGTCGTTTCCTCCTCCTCCTCCTGCGCAGGGCAGAGCCTGGACATTAAAAC 747
      | : :: : : : : | : : : : : : : :
Db      481 EASFMSVNESCYKYGTLDLSKNSIFFVKSSDFQHLNFLKCLNLSGNLISQTLNGSEFQP 540
```

Qy 748 ATGCCCTGCCTGAAGCCGCTTGCTGCTTCTCACTGATTTCTGCTCTCCCCCTTCCTTGACT 807
 : : : : : : : | : : : : : : : | : : : : : : :
 Db 541 LAELRYLDFSNNRLDLLHSTAFAEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKVLOK 600

 Qy 808 CGCCCACCACCTGTCCTGTGTAGATGGAGAAGGCTCGGAGAGTGGGGGTGCTGGGGGCAC 867
 : : : : : : : : : : : : : | : : : : : : :
 Db 601 LMMNDNDISSSTSRTMESESLRLEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKN 660

 Qy 868 AAAATGGAATGAACACTGCTGAAGGAATGCAGGGTTC 904
 : : : : : : : : : : : : : | : : : : : : :
 Db 661 SLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQC 697

Search completed: April 29, 2004, 21:08:49
 Job time : 1532.81 secs